AN EXAMINATION OF THE EFFECTS OF VIDEOPHONES ON DRIVING AND CONVERSATION PERFORMANCE

A thesis

submitted in fulfilment

of the requirements for the degree

of

Master of Arts in Psychology

at

The University of Waikato

by

KATHRYN JULIE MACKENZIE

The University of Waikato

2011

Abstract

Research has conclusively shown that cell phones have a detrimental effect on driving performance. In an attempt to understand why, a handful of researchers have investigated the differences between cell phone and passenger conversations, with several of these studies revealing that the distraction caused by concurrent cell phone conversations noticeably outweighs that imposed by passenger conversations. One study suggested that the availability of visual cues during a passenger conversation may be an important factor contributing to this reduced level of distraction. The focus of this research project was to test whether providing drivers and remote conversers with access to visual cues via a videophone would result in improved driving performance when compared to a concurrent cell phone conversation. An initial experiment, in which 24 drivers encountered five hazards on a simulated road while conversing with a passenger, cell phone caller, videophone caller, or driving without conversation, resulted in driving behaviour that did not appear to be an accurate representation of real-world driving behaviour, which resulted in the early termination of this experiment. A second revised experiment, in which novice and practiced drivers drove a shortened version of the simulated road once under each of the aforementioned conversation conditions, produced more normal behaviour but failed to reveal any significant differences in driving or conversation performance as a result of concurrent videophone conversation compared to cell phone conversation. However, the results did reveal a number of other findings that may aid in understanding the distracting effects of cell phones, one of which was that remote conversations may result in an overestimation or underestimation of the correct driving response depending on the nature of the driving situation.

Acknowledgements

There are numerous people who have played an invaluable role in helping me achieve this Masters.

First and foremost, I would like to thank my first supervisor Sam Charlton, who has helped me every step of the way; so often going above and beyond the call of duty to see me and this project through. Thank you so much for your time, expertise, support and understanding.

My second supervisor, Robert Isler – you have also been a tremendous help. Thank you for all your advice and support along the way.

Rob Bakker – thank you for building whatever I asked you to build and for helping me solve the various simulator problems I encountered. Although, I pretty much just gave you a reason to make more toys and get paid for it, so you should be thanking me ;-)

Dad – you built something that did in 30 seconds what took me 4 weeks to do manually – legend. Just wished I found out what you could do at the very beginning instead of half way through. Still saved me a massive amount of time in the second half so thanks heaps xx

To my friends and family and everyone else who encouraged me along the way – thank you so much. I am so stoked to have such an awesome group of friends and I could not have done this without you.

To all my participants, thank you so much for your time and enthusiasm. It made the whole thing possible.

Lastly, I would like to thank the organisations who have financially contributed to this project through scholarships – the University of Waikato, the Faculty of Arts and Social Sciences, the Waikato Police, and the New Zealand Federation of Graduate Women – Waikato Graduate Women Charitable Trust. The scholarship funds were greatly appreciated.

Table of Contents

Abstract ii

Acknowledgements iii

Table of Contents v

List of Figures viii

Introduction 1

The problem 1

Identifying the source: physical versus cognitive distraction 3

Differences and similarities between passenger and cell phone conversations 7

Research indicating no difference between cell phone and passenger conversations 11

Research finding differences between passenger and cell phone conversations 25

The current study – outline, research questions, and hypotheses.................. 36

Method – Experiment 1 39

Participants... 39

Apparatus 39

The simulated driving scenario 41

Procedure 48

Data collection and analysis 50

Results – Experiment 1.. 52

Driving Performance.. 52

Speed 52

Reaction time and time-to-collision 56

Crashes.. 60

Rest area task... 61

Conversation Measures 61
Utterance length 63
Pauses 65
Number of words 65
Situation awareness utterances 66
Difficulty and interference ratings 68
Discussion – Experiment 1 74
Introduction – Experiment 2 81
Method – Experiment 2 88
Participants 88
The simulation scenario 89
Procedure 91
Data collection and analysis 92
Results – Experiment 2 93
Driving performance 93
Speed 93
Reaction time and time-to-collision 103
Crashes 109
Strategic task: rest area 109
Conversation performance 110
Utterance length 111
Situation awareness utterances 113
Total number of words 115
Ratings 116
Driving difficulty 116
Conversation difficulty 117
Conversation interference 119

Discussion – Experiment 2 122
The overall effects of videophones 122
Conversation modulation in videophone conversations 124
Additional findings of importance 126
The relationship between speed and utterance length 134
The influence of hazard type on speed 136
The story told by a lack of results 138
Summary 141
Limitations 142
Directions for future research 145
References 146
Appendices 152
Appendix A: Participant Information Sheet – Experiment 1 152
Appendix B: Habit questionnaire used in Experiments 1 and 2 154
Appendix C: Participant instructions – Experiment 1 156
Appendix D: Conversation ideas card used in Experiments 1 and 2 160
Appendix E: Participant information sheet – Experiment 2 162
Appendix F: Participant instructions – Experiment 2 164
Appendix G: Table showing counterbalanced order for participant trials used in Experiment 2 167

List of Figures

Figure 1. The University of Waikato driving simulator 42

Figure 2. Internal view of the University of Waikato driving simulator, showing set up for Videophone condition 42

Figure 3. Map of Experiment 1 simulation scenario 46

Figure 4. Hazard 1 – busy intersection with turning police car 47

Figure 5. Hazard 2 – previously obscured parked car pulling onto the road 47

Figure 6. Hazard 3 – one lane bridge 47

Figure 7. Hazard 4 – road works 47

Figure 8. Hazard 5 – slip 47

Figure 9. Rest area 47

Figure 10. Mean speeds at Hazard 1 – busy intersection 53

Figure 11. Mean speeds at Hazard 2 – parked car entering traffic 53

Figure 12. Mean speeds at Hazard 3 – one lane bridge 53

Figure 13. Mean speeds at Hazard 4 – road works...54

Figure 14. Mean speeds at Hazard 5 – slip 54

Figure 15. Reaction times and times-to-collision for Hazards 2-5 58

Figure 16. Mean utterance length of drivers 63

Figure 17. Mean utterance length of conversers..64

Figure 18. Mean number of pauses for drivers and conversers 65

Figure 19. Mean total number of words spoken by participant dyads 66

Figure 20. Mean percent of situation awareness utterances for drivers and conversers 67

Figure 21. Drivers' ratings of driving difficulty...69

Figure 22. Drivers' ratings of conversation difficulty..70

Figure 23. Drivers' ratings of conversation interference 71

Figure 24. Conversers' ratings of conversation difficulty 72
Figure 25. Conversers' ratings of conversation interference 73
Figure 26. Map of Experiment 2 simulation scenario 91
Figure 27. Mean speeds of novice drivers at Hazard 1 95
Figure 28. Mean speeds of practiced drivers at Hazard 1 95
Figure 29. Mean speeds of novice drivers at Hazard 2 98
Figure 30. Mean speeds of practiced drivers at Hazard 2 99
Figure 31. Mean speeds of novice drivers at Hazard 3 101
Figure 32. Mean speeds of practiced drivers at Hazard 3 102
Figure 33. Mean reactions times at Hazard 2 106
Figure 34. Mean time-to-collision values for Hazard 2 106
Figure 35. Mean reaction times at Hazard 3..108
Figure 36. Mean time-to-collision values for Hazard 3 108
Figure 37. Mean utterance length of drivers 112
Figure 38. Mean utterance length of conversers...112
Figure 39. Mean percent of SA utterances – drivers 114
Figure 40. Mean percent of SA utterances – conversers 114
Figure 41. Mean total number of words spoken by participant dyads 115
Figure 42. Drivers' ratings of driving difficulty 117
Figure 43. Drivers' and conversers' ratings of conversation difficulty 118
Figure 44. Drivers' and conversers' ratings of conversation interference 120

Introduction

The problem

The effect of concurrent phone use on driving performance has been a topic of interest for over 40 years. In the first article published on the subject, Brown, Tickner, and Simmonds (1969) reported that conversing on a phone while driving affected gap selection, response times and driving speed. Since then, an extensive amount of research has been conducted in the area, with a 2005 literature review examining 78 articles pertaining to the topic (Svenson & Patten, 2005). More recently, the volume of research was estimated to be about a metre in width (Caird, Willness, Steel, & Scialfa, 2008). This research spans a wide variety of settings, and includes laboratory tests (e.g., Wood et al., 2006), high-fidelity driving simulator studies (e.g., Charlton, 2009; Cooper & Strayer, 2008), on-road field tests in actual vehicles (e.g., Crundall, Bains, Chapman, & Underwood, 2005), and observational studies (e.g., Rosenbloom, 2006). The tasks required of participants have included a variety of verbal information-processing tasks (e.g., Gugerty, Rakauskas, & Brooks, 2004), mental arithmetic tasks (e.g., Horrey, Lesch, & Garabet, 2008), and naturalistic conversation (e.g., Strayer, Drews, & Johnston, 2003), among others.

Measures of driving performance have also varied between studies. Response time has been one of the most commonly used measures (e.g., Alm & Nilsson, 1995; Beede & Kass, 2006; Charlton, 2009; Consiglio, Driscoll, Witte, & Berg, 2003; Hancock, Lesch, & Simmons, 2003; Strayer & Johnston, 2001), and has also produced the most consistent results. Meta-analyses have revealed a consistent increase in response time for braking as a result of concurrent cell phone use (Caird et al., 2008; Horrey & Wickens, 2006). Other measures have produced more contradictory effects. For example, Alm & Nilsson (1994) found that concurrent cell phone use reduced speed, whereas Rosenbloom's (2006) observational study found that short conversations had no effect on speed whilst long conversations led to an increase in speed in a sample of Israeli drivers. This pattern is repeated when headway is used as a measure. Strayer et al. (2003) found that drivers increased the distance between themselves and a lead car whist conversing on a cell phone, whereas Rosenbloom (2006) observed a reduction in headway amongst cell phone-using drivers. Lane-keeping results have also been inconsistent, with several studies identifying a reduction in lane-keeping variability (e.g., Becic et al., 2010; Beede & Kass, 2006; Brookhuis, de Vries, & de Waard, 1991; Rakauskas, Gugerty, & Ward, 2004), but meta-analyses have been unable to reconcile results in this area (Caird et al., 2008; Horrey & Wickens, 2006).

Any doubts these contradictions may raise about the effects of concurrent cell phone use on driving performance can be put to rest when epidemiological research is taken into account. A widely cited epidemiological study found that using a cell phone while driving leads to a 400% increase in crash risk, suggesting a level of impairment similar to that displayed by drunk drivers (Redelmeier & Tibshirani, 1997). A complementary study by Strayer, Drews, and Crouch (2006) compared the effects of cell phone use and alcohol in a laboratory setting and found cell phone use resulted in impairment worse than that displayed by drivers with a 0.08% blood alcohol level. Thus, despite some differences in methodologies and direction of changes in observed measures of performance, there is general consensus that using a cell phone while driving has an adverse effect on driving performance.

Identifying the source: physical versus cognitive distraction

In an effort to determine the exact cause of these deleterious effects, several researchers have explored the role of physical distraction by comparing handheld and hands-free technology. In a study by Brookhuis et al. (1991), 12 subjects drove an instrumented vehicle on actual roads for an hour a day, five days a week, for three weeks. Each drive included a range of different road types and traffic conditions, as well as periods of silence and periods of cell phone conversation. During the times of conversation, half of the participants used a handheld phone and the other half used a hands-free model. Brookhuis et al. found drivers displayed significantly poorer control of the vehicle during the dialling phase of the handheld cell phone task than at any other time. However, they also found that both types of cell phone conversation caused a significant increase in both subjective and objective measures of workload when compared to driving without conversation, with no significant difference between phone types for these workload ratings. A similar pattern was observed by Tornros and Bolling (2005), who found no significant difference between hands-free and handheld phones for subjective ratings of mental effort. Like Brookhuis et al., they also noted poorer vehicle control by participants during dialling. Interestingly, they noticed no effect of phone type, even though dialling in the hands-free mode did not require any physical manipulation of the phone. They also found that while lane-position variability increased during the dialling phase for both phone types, the opposite effect was noticed during the conversation phase.

Several other studies have demonstrated that handheld and hands-free cell phones cause similar levels of impairment in various factors related to driving performance, such as tracking (Strayer & Johnston, 2001) and peripheral detection (Patten, Kircher, Ostlund, & Nilsson, 2004). An Australian case-crossover study found no effect of phone type on serious crash risk (McEvoy et al., 2005). It must therefore be concluded that although the manual handling of a cell phone may cause some physical distraction, an even greater proportion of the distraction associated with cell phone use while driving must be attributed to the cognitive demands imposed by the conversation itself.

It is still unclear exactly what aspect of cell phone conversation contributes most to this cognitive distraction. The effects of conversation difficulty have been explored by several researchers, and once again, results have been mixed. Patten et al. (2004) concluded that conversation difficulty did affect driving performance, based on the results of their study in which participants fielded a number of cellular phone calls while driving on a Dutch motorway. The difficulty level of the ensuing conversations was manipulated, with the simple conversation condition involving an easy addition task, and the difficult conversation condition consisting of a mathematical task incorporating both addition and memory. Workload was measured indirectly through a peripheral detection task, and it was found that conversation difficulty had a significant effect on performance of this task. With the exception of speed, driving performance itself was not measured, so it is impossible to determine whether this increase in workload would necessarily translate into noticeable driving performance decrements. The authors did report an effect of phone type on speed, with those in the handheld condition driving slower, but no effect of conversation difficulty was noticed.

A study by McKnight and McKnight (1993) revealed an age effect in regards to conversation difficulty. Participants drove a simulator through pre-recorded highway scenarios and their response rate to events in these scenarios was calculated. Responses were determined by steering movements, signal use, acceleration, and deceleration. Conversation tasks included placing a call, casual conversation with an experimenter, and intense conversation, which involved solving mathematical problems. While all conversation-related tasks produced significant changes in response rate to highway events for all drivers, the intense conversation condition produced the largest effect in drivers under 50. Drivers over 50 displayed greater performance decrements than their younger counterparts while placing a call and in the casual conversation condition; however their response rate in the intense condition matched that of the younger drivers. McKnight and McKnight determined that conversation difficulty may affect driving performance in younger drivers, but due to a diminished capacity for task sharing, this effect is less noticeable in older drivers.

Harbluk, Noy, Trbovich, and Eizenman (2007) also concluded that secondary task difficulty affects driving performance. In this on-road experiment, participants drove an instrumented car on a 4-lane city road while engaging in hands-free cell phone conversations consisting of either simple addition problems (easy task) or more complicated mathematical problems (difficult task).

The authors found that drivers spent more time looking straight ahead than in their periphery when engaging in a difficult conversation as opposed to no conversation, and also made fewer glances towards their mirrors and instruments. They also displayed less visual scanning behaviour at intersections during the difficult task compared to the no conversation condition, and an increase in hard braking was observed.

However, it has been argued that mathematical tasks such as the ones used in the above experiments do not accurately represent the conditions drivers experience when engaging in real-life cell phone conversations. Shinar, Tractinsky, and Compton (2005) compared emotional conversations with a mathematical task and found that the mathematical task caused a greater level of distraction than the emotional conversations. (The topics for the emotional conversations were based on information gathered during prior interviews with participants). This provides evidence for the idea that while an increase in cognitive workload may produce greater deficits in driving performance, changes in content within naturalistic conversations may not be sufficient to produce this increased workload and the associated effects.

This idea was further supported by Rakauskas et al. (2004), who also investigated conversation difficulty, but utilised naturalistic conversation for both the easy and difficult conversation conditions. The driving task involved participants driving on a simulated rural road that included lights, stop signs, and traffic hazards. The conversation task involved participants answering pre-determined questions, with each question being classified as either easy or difficult based on the results of pilot studies. The authors found no effect of conversation difficulty on measures of workload, speed maintenance, lane-position maintenance and crash avoidance. The null workload finding was particularly interesting given that pilot studies had revealed a significant difference in workload ratings for the easy and difficult questions when they were answered by participants who were not driving. It may be that when driving is incorporated as a dual-task, the increase in workload imposed by the driving task renders the driver less sensitive to detecting changes in conversation difficulty.

Differences and similarities between passenger and cell phone conversations

As another means of understanding exactly which aspect of cell phone conversations make them so detrimental to driving performance, several researchers have turned to comparisons with passenger conversations. As Irwin and Chekaluk (2006) point out, conversing while driving had never really been considered a problem until cell phones were introduced. Therefore, it could be presumed that passenger and cell phone conversations are fundamentally different at some level. Authors differ as to where they believe these differences are hidden, with some maintaining the previously-discussed conversation difficulty hypothesis (Nunes & Recarte, 2002); while others suggest they possibly lie in conversation form and content (e.g., Charlton, 2009; Drews, Pasupathi, & Strayer, 2008); or in access to non-verbal cues (Hunton & Rose, 2005).

While a significant amount of research has been conducted in the wider area of the effects of cell phones on driving, research comparing passenger and cell phone conversations is scarce. The 12 articles reviewed here form a large majority of the available research pertaining directly to this topic. Results have so far been mixed. Of these 12, seven suggest conversing on a cell phone is no more detrimental than talking to a passenger (Amado & Ulupinar, 2005; Consiglio et al., 2003; Gugerty et al., 2004; Irwin & Chekaluk, 2006; Nunes & Recarte, 2002; Rivardo, Pacella, & Klein, 2008; Waugh et al., 2000). The remaining five have revealed significant differences between the effects of passenger and cell phone conversations on a range of measures pertaining to driving (Charlton, 2009; Crundall et al., 2005; Drews et al., 2008; Hunton & Rose, 2005; McEvoy, Stevenson, & Woodward, 2007). The 2007 article by McEvoy et al. is based on two complementary passenger and mobile phone epidemiological case-crossover studies, conducted simultaneously. The results of the cell phone study were also published independently by McEvoy et al. (2005), but for the purposes of this review, only the 2007 comparison article will be examined. Some may consider it presumptuous to affirm the idea that passenger and cell phone conversations are different at some fundamental level when over half of the research conducted to date has revealed no difference between the effects of the two types of conversation on driving performance. However, when the quality, as opposed to quantity, of the research is taken in account, some very clear patterns emerge and a very different overview is obtained.

Of the seven articles asserting that passenger and cell phone conversations have the same effect, most were conducted in low-fidelity simulators, with only one being an on-road field test (Nunes & Recarte, 2002) and one other involving the use of a closed test track (Waugh et al., 2000). Of the five revealing differences, one was based on two large-scale epidemiological studies as mentioned previously (McEvoy et al., 2005; McEvoy et al., 2007), another was an on-road field test (Crundall et al., 2005) and the remaining three were conducted in high-fidelity simulators (Charlton, 2009; Drews et al., 2008; Hunton & Rose, 2005). The driving tasks required of participants in these studies were very similar to what drivers may encounter in real-world driving, whereas the experiments producing results showing little or no difference usually involved participants performing only one aspect of a the driving task, such as braking (Consiglio et al., 2003) or hazard identification and recall (Gugerty et al., 2004). Some were based purely on cognitive laboratory tasks (e.g., Amado & Ulupinar, 2005). Thus it could be said that in general, the experiments that revealed differences between the effects of the two types of conversation had higher ecological validity, potentially giving greater weight to their results.

It is also interesting to note that the methodological design of the passenger and cell phone conditions differed considerably between these two groups of studies. All of the studies that found significant differences between the effects of cell phone and passenger conversations had the passenger seated next to the driver, just as a passenger would be in a real car. They also all used actual phones to create the cell phone condition. In the experiments showing no difference, both the passenger and cell phone conditions were represented in a variety of ways. Some saw the passenger sitting in the back seat (Consiglio et al., 2003) or even posing as another student turning up to the study, striking up an apparently impromptu conversation with the participant as they completed the task (Irwin & Chekaluk, 2006). Cell phone conditions (perhaps better referred to as remote converser conditions) ranged from the use of an actual mobile phone (e.g., Nunes & Recarte, 2002; Waugh et al., 2000), a loudspeaker (Amado &

Ulupinar, 2005), through to the conversation partner sitting beside the driver like a passenger but being unable to see the driving scenario (Gugerty et al., 2004; Rivardo et al., 2008). These methodological differences are also likely to have affected the ecological validity of the research.

Further differences can be found in the type of conversation task chosen, and whether or not conversation was free to vary or modulate based on driving demands. This idea of conversation modulation, in which the driver and conversation partner alter the form and content of their conversation in response to the real-time demands of the driving situation, has been suggested by several researchers as the key feature differentiating passenger and cell phone conversations (e.g., Charlton, 2009, Drews et al., 2008). If so, the decision to allow or avoid conversation modulation in an experiment may significantly affect the outcome. Not surprisingly, the majority of the experiments revealing differences between the effects of passenger and cell phone conversation allowed conversation modulation to occur, while most of those not finding differences did not. Rivardo et al.'s study (2008) was the only one of the seven not finding differences to allow free-flowing conversation between two naïve participants to occur, but unfortunately the equipment used to record the experimental sessions did not allow for post-hoc discourse analysis and the presence or absence of conversation modulation was unable to be determined.

The patterns discussed above present several possible reasons for the discrepancies in the research findings to date. However, as is the case with most research, each of these experiments was designed to achieve certain objectives and test certain hypotheses. Thus, in order to truly understand each individual set of results and the body of research as a whole, it is necessary to briefly examine each study individually, in light of the unique combination of factors through which it was produced. The studies reporting no difference between passenger and cell phone conversations will be reviewed first.

Research indicating no difference between cell phone and passenger conversations

In 2000, Waugh et al. analysed the driving and conversation performance of 12 participants who drove a closed-loop serpentine test track in their own cars while completing memory and problem-solving tasks verbally. The verbal tasks were presented by an experimenter either posing as a passenger or over a mobile phone. Driving performance was measured by lap completion times and number of errors (cones hit). Conversation performance was assessed by response times and accuracy scores in the verbal tasks. Workload was also measured using the Subjective Workload Assessment Technique (SWAT), which was administered every 20 seconds. In regards to driving performance, Waugh et al. found that lap times increased from baseline to passenger, and then further still from passenger to mobile phone. However, error rate remained constant throughout. Response times for the verbal task followed the same pattern, yet once again there was no significant effect between conditions for error rate. Mental workload was lowest for the drive-only baseline, and this was significantly different from all other conditions. Workload increased steadily through the verbal task baseline, passenger, and mobile phone conditions. Workload ratings for the mobile phone condition were significantly different from those reported during the baseline verbal task condition, but the passenger condition was not significantly different from either.

It appears that in this case, the nature of the driving and conversation tasks may have prevented an effect of conversation condition being seen in driving performance. While the driving task used in this experiment required constant activity on the part of the driver (steering, acceleration, braking), it required very little situation awareness (see Endsley, 1995). Real roads present dynamic driving scenarios that require accurate interpretation and response selection on the part of the driver in order to be negotiated successfully. In this case, drivers were able to maintain performance by merely reducing their speed, but had this experiment involved unexpected hazards, a different result may have been obtained.

A more naturalistic conversation task may also have produced a different outcome. Drivers slowed their performance in the conversation task which allowed them to maintain a constant level of accuracy. However, this may not be possible in real cell phone conversations as these are subject to being paced by someone who is unaware of the driving scenario. In fact, as will be discussed in context later, Drews et al. (2008) found that even in situations where a driver attempted to modulate a cell phone conversation, this modulation was not reciprocated by a cell phone converser. In summary, methodological issues are likely responsible for the lack of a difference between passenger and cell phone conversation being noticed in this case.

Three years later, Consiglio et al. (2003) investigated how the effects of cell phones on braking responses compared to other potential distracters, including passenger conversation. Consiglio et al. felt that cell phones had been given undue attention as a distracter. In this study, 22 participants were told to attend to a red lamp in front of them, and release the accelerator and depress the brake pedal when the lamp came on. The lamp was positioned so as to simulate a lead car braking 12 metres in front. Response times were compared across five conditions: control, listening to music (radio condition), passenger conversation, handheld phone, and hands free. Conversation involved the experimenter asking the participant scripted questions about a range of topics such as their interests, studies, and family. Similar to Waugh et al.'s (2000) results, it was found that all conversations resulted in significantly slower reaction times but no difference between conversation types was noticed. The impairment in the radio condition was not significantly different from baseline. Although the conversation task in this experiment was relatively naturalistic, it did not allow for conversation modulation. The nature of the driving task used in this experiment may have meant that this omission had an even greater effect than in other studies. This particular driving task consisted of multiple trials, with the lamp being activated once per trial, 10-20 seconds after the trial began. The trial ended after the lamp activation and response. Therefore, the driver was either anticipating the need to respond or initiating a response for the entirety of each trial. Times of high-demand driving such as this are the exact times during which an aware passenger would likely suppress their conversation. It is therefore possible that given the option, a naïve conversation partner would have chosen not to speak at all in this scenario, just as a passenger might remain silent while a driver passed a school as the children were being let out, for example. By forcing participants to engage in conversation during a time when it would have likely been suppressed in a real driving situation, any differences between passenger and cell phone conversations remained hidden. Another potential contributing factor was that the driver was instructed to not look at the passenger during the conversation, making an apparently naturalistic conversation appear slightly less natural.

Irwin and Chekaluk (2006) also wanted to see how the distraction caused by mobile phones compared to other potential distracters. The experiment involved four conditions, two of which were handheld cell phone and passenger. Rather than including a radio condition however, they chose instead to assess the distraction caused by merely listening to a conversation between two passengers. Given the focus on passengers in this experiment, the control condition was essentially a silent passenger condition, in which someone was present in the room but did not converse with the driver. Eighty participants drove a simulated road through both urban and rural areas under one of the above conditions. On the drive, participants encountered traffic lights, speed and hazard signs, as well as pedestrians. The number of crashes and number of tickets (given for speeding and running traffic lights) was used to assess performance. In the cell phone condition, an experimenter called the driver on their phone and asked questions relating to their participation in the study. In the two-passenger condition, two confederates, posing as other students waiting to take part in the study, sat in the same room as the driver and performed a scripted general conversation about coursework. In the single-passenger condition, a confederate followed a similar script but this time engaged the driver with questions about their studies.

No significant effect of condition was found in any of the measures when the groups were analysed as a whole. However, a significant effect of gender was noticed. Females produced the worst performance in the mobile phone condition, but performed significantly better than control in the single and two-passenger condition. Males, on the other hand, performed worst in the control condition, and best in the cell phone condition.

This unusual result could be reflective of the age of the participants and also the design of each distracter condition. Younger drivers have been found to commit more driving violations than older drivers, and this effect is more pronounced in males than females (Harré, Field, & Kirkwood, 1996; Jonah, 1990). Young males' driving behaviour is also likely to be worse in the presence of male passengers (Simons-Morton, Lerner, & Singer, 2005). It was not specified whether the confederates were male or female, but even so, it may be that in each of the three passenger conditions the male participants were pushing the limits of the simulator, and perhaps not even treating the experiment as real. The better performance in the cell phone condition could be explained by the lack of an additional person in the room, or it may be that this particular condition caused such an increase in mental demand that the participants no longer had sufficient cognitive resources available to focus on intentionally altering their driving performance. Therefore, it appears that once again, methodological issues prevented these two studies from definitively demonstrating how cell phone conversation compares to other potential distracters, including passenger conversation.

In 2005, Amado and Ulupinar set out to test the effect of conversation type (in-vehicle vs. remote) and difficulty on attention and peripheral detection. To achieve this, 48 undergraduate students completed two tasks – the Cognitrone and the Peripheral Detection and Dual-processing task (PDDpT) – while answering evenly-paced general knowledge and mathematical questions. The Cognitrone was a computer based matching task for which the number of correct responses and response time were used as a gauge of attention. The PDDpT saw participants steering a basic simulator down a straight road, taking care to keep within their lane and avoid oncoming traffic, while detecting LED lights that appeared on either the right, left, or both sides simultaneously. Performance on this task was based on number of correct responses, response time, and the number of road line contacts and vehicle collisions. Amado and Ulupinar found that while conversation degraded performance on both tasks, there was no effect of conversation type.

Conversation difficulty was also manipulated in the study and produced results suggesting passenger conversation is in fact more demanding than cell phone conversation. The questions used in the conversation task were rated as either easy or complex by a separate group of pilot participants. Fewer complex questions were answered in the in-vehicle conversation condition during completion of both the Cognitrone and PDDpT, despite task performance being similar across both conversation conditions. It is possible, and is suggested by the authors, that this was a result of performance anxiety produced by the experimenter's presence in the room (note the experimenter was holding a stop watch to pace the questions accurately). If so, this suggests that more natural, anxiety-reducing conversation tasks are desirable to allow for accurate detection of any effects of conversation type. It is also possible that participants prioritised the conversation lower during the in-vehicle condition than during the cell-phone condition. In reality, drivers and passengers usually have shared goals other than the conversation itself, such as reaching a common destination. Conversation serves to pass the time and is socially appropriate. However, interaction involving a cell phone is usually centred on the conversation itself. It is rare to call someone and neither party talk; however periods of silence would be perfectly fine in a passenger situation. Consequently, drivers may focus more on a cell phone conversation because it must be maintained at a certain level to be considered normal. Therefore, it may have been that drivers felt more comfortable missing questions in the in-vehicle condition, given that the experimenter was present and could see the driver was completing another task.

One major weakness of this study is that none of the participants were licensed drivers. While this may have had little effect on performance in the Cognitrone task, the PDDpT included a driving element. Given that driving is a skill that requires a considerable amount of practice until it can be performed competently and automatically, this potentially throws the entire experiment into question. Amado & Ulupinar (2005) defend their choice of task by citing research that shows that performance in these tasks is significantly related to accidents and driving experience (Amado, Koyuncu, & Kacaroglu, 2004, cited in Amado & Ulupinar, 2005). Yet if performance on these tasks is related to driving experience, performance by a sample of non-licensed participants should be expected to be poor as driving experience is likely to be considerably low or non-existent.

Nunes and Recarte (2002) also chose to investigate the role of conversation complexity on driving performance. They hypothesised that distraction associated with cell phone use while driving was purely a product of conversation complexity, and that if a passenger conversation (or even one's own thoughts) became complex enough, a similar level of distraction would ensue. To test this hypothesis, they conducted four experiments. All experiments were conducted on-road in an instrumented car, with an experimenter in the car at all times. This instantly raises questions about the validity of the study as it means that during all cell phone conditions, a passenger was present. There is no mention as to whether the experimenter was seated in the front or back seat; however it is possible that their mere presence could have affected results, regardless of position.

Nunes and Recarte (2002) chose to use ocular measures as the main gauge of distraction and performance. Fixation duration, pupil size, spatial gaze variability, blink rate and duration, as well as glance rates for mirrors and speedometer were all calculated. Driving performance was measured by speed only. During Experiment 1, six participants completed 21 cognitive tasks whilst driving, two of which were phone tasks and one being a casual conversation with the experimenter. Surprisingly, the phone tasks were rated the easiest of all tasks and produced the smallest changes in visual behaviour. The data also showed no differences between passenger and phone conversations. The authors conceded the phone tasks may have caused cognitive impairment that was undetected by visual measures, so they conducted a second experiment of similar design, but which incorporated a task that required detection, discrimination and response selection. However, whilst all other cognitive tasks, including passenger conversation, led to changes in visual behaviour and negatively affected performance on the detection task, the phone conversation did not. The authors did not address the difference between passenger and cell phone conversations found here, but rather ignored it completely, placing emphasis instead on the difference found between the phone task and other apparently harder cognitive tasks.

The effects of conversation complexity were investigated in the third experiment. Complexity was manipulated by administering two difficult cognitive tasks by both phone and in person, and comparing them to the apparently low-demand naturalistic conversations used in experiments 1 and 2. It was found that the difficult cognitive tasks produced changes in visual behaviour whereas the naturalistic conversations did not. No effect of conversation type (passenger versus cell phone) was found. In Experiment 4, detection task performance was the main focus, and once again, no effect of conversation type was noticed.

One weakness of this study is that all the experiments could be described as extremely "busy." In Experiment 1, participants completed 21 cognitive tasks; in Experiment 2, this number was reduced to 10. Experiment 3 had only three tasks and in Experiment 4 the number increased to 13. Little mention was given to the nature of most of these tasks, which prevents readers from judging for themselves whether any of these tasks may have negatively interfered with any of the others. The logic of these numbers or the justification for each task was also never discussed. This ‘busyness’ also meant little time was allowed for each task, with each of the phone conversations in Experiment 1 lasting only two minutes. Participant numbers were also small, with only six being used in Experiments 1 and 2, and 12 being used in Experiments 3 and 4. As such, the claims that arose from this research were based on data collected from a small number of participants over a short period of time.

Another potential cause for interference that the authors did not address in enough detail was the instruction given to all participants that they need not answer the phone if they felt unsafe. Given that this was an on-road experiment, this is a necessary precaution. However, in order to validate the results, the number of times this occurred should have been reported. Also, the majority of cognitive tasks were administered by an experimenter in the vehicle. There is no reference to the pacing of these tasks and given that safety was obviously a concern for the researchers, it is possible that the experimenter timed the administration of these tasks with periods of low-demand driving, in essence modulating the conversation. Had participants in this study been forced to complete tasks at non-optimal times, a different set of results may have been obtained.

Several of the studies described so far were conducted with the intent of demonstrating that the effects of cell phone conversations are no worse than those of passenger conversations. However, there were two studies that, despite the authors hypothesising a difference between the effects of passenger and cell phone conversations, still revealed no difference between the two conversation conditions. Gugerty et al. (2004) produced one such study. Twenty-nine participant pairs completed the experiment, with one participant acting as driver and the other as conversation partner. Driving participants were shown two blocks of 35 scenes. Each scene lasted 18-35 seconds long. At the end of a scene, drivers were probed about it in one of three ways. Location-recall probes involved the driver indicating where the other cars were once the screen disappeared. Performance probes involved the driver needing to make a response to a hazard. Scene-interpretation probes involved the driver making judgements about other cars and their probable movements. One block of 35 scenes was completed concurrently with a verbal task, the other block without it. The order of this was counterbalanced. The verbal task involved the teammates taking turns to say words, with each new word having to start with the last letter of the previous word. The experimenter gave the first word to start the task and the non-driver went next. For the in-car passenger condition, the conversation partner sat next to the driver, looking at the screen. For the remote condition, the conversation partner sat 1.5m away behind a screen, which blocked their view of both screen and driver without impairing audio quality.

Gugerty et al.'s (2004) original hypothesis was that conversers should have a slower speech rate in the in-person condition than in the remote condition, with the reverse being true for drivers. This hypothesis was based on the idea that conversers would initiate conversation modulation during the in-person condition as they would be aware of the task demands; whereas during the remote condition, drivers would assume this responsibility as the converser could not be expected to without being able to see the driving scenario.

Instead, Gugerty et al. (2004) observed a greater slowing of speech by both parties in the remote condition than during the in-person condition. They also found that situation awareness in the spatial task was comparably degraded by both types of conversation, meaning poorer performance in many skills pertinent to safe driving, including hazard detection and response. One factor that may have potentially influenced the results was the type of conversation task chosen for the experiment. While the literature concerning the effects of cell phone conversation on driving performance show researchers have employed a range of different tasks, this one may perhaps bear the least resemblance to actual conversation.

The authors suggested that their driving participants may likely have prioritised the driving task over the verbal task, thus as the verbal task became harder (i.e. in the remote condition), they allocated less attention to the verbal task in order to maintain their performance in the driving task. In order to control for this, they conducted a second experiment, in which participants were offered monetary rewards for good performance in both the driving and verbal task. The verbal task was altered so that the conversers were presented with words to use, meaning only the driver had to think of words fitting the last-letter rule. All conversers were unable to see the drivers but some conversers were shown the highway scenarios on a separate screen. The converser's rewards were based on the driver's performance so as to simulate a real-life passenger or caller wanting a driver to perform well to stay safe.

Driving and conversation appeared to negatively affect each other, with lower situation awareness being displayed by drivers while conversing than during baseline, and poorer verbal performance being displayed by drivers than conversers. However, once again no difference was found between conversation types. Gugerty et al. (2004) argue that in this particular case it was a result of the conversers in neither condition being able to see the driver's face, an idea that will be explored in more detail later.

Rivardo et al. (2008) also set out to test the theory that passenger conversations are safer than cell phone conversations due to the passenger's ability to modulate the conversation based on driving demands. Instead they found that simulated driving performance was actually worse when talking to a passenger. However, this could possibly be attributed to their experimental design. Seventy-nine student pairs shared holiday memories while completing three driving trials under different conversational conditions each time – a no-conversation control, a normal passenger condition and a blind passenger condition. The normal passenger condition designed by Rivardo et al. resembled what a driver may experience in a real-life driving scenario and was similar to those employed by most other researchers investigating differences between passenger and cell phone conversations. The conversation partner was seated next to the driver and could see the driving scenario as it unfolded. Yet in place of a cell phone condition with a proximally distant remote caller, Rivardo et al. opted for a blind passenger condition, in which conditions were exactly the same as their normal passenger condition except the passenger was facing backwards, unable to see the screen. This design was chosen to control for possible interference as a result of transmission, however it may have created a set of experimental conditions that did not accurately represent conditions experienced when engaging in a real cell phone conversation while driving.

Rivardo et al. (2008) suggest several alternate explanations for their unusual findings. The first is that the normal passenger condition may have elicited more risky driving behaviours, given the young age of their sample and the previously-mentioned finding that risky driving behaviours are more likely to occur in teenaged drivers when a male passenger is present (Simons-Morton et al., 2005). They also suggested that the conversation task they used (getting participants to talk about their favourite vacations) was not difficult enough to elicit any noticeable effect on driving performance. However, Charlton's 2009 study saw participants engaging in naturalistic conversation with no content restrictions and still revealed significant differences in driving performance between those in the passenger and cell phone groups. Thus it seems more likely that Rivardo et al. created a conversation condition that emphasised the positive features of both passenger and cell phone conversations while minimising the negative features. One way of testing for this would have been to analyse the conversations for evidence of modulation. Unfortunately in this case, the equipment used to record the session did not allow for post-hoc discourse analysis and the presence or absence of conversation modulation was unable to be determined.

The authors reported that participants may have felt pressured to perform well, as prizes were awarded to the three best-performing teams. This pressure may have contributed to poorer performance in the normal passenger condition, especially if participants felt their performance was being scrutinised by the passenger. They may have also felt a need to attend to the passenger from time to time. It is possible that in the blind passenger condition, this feeling of scrutiny and the need to pay attention to the passenger may have both been substantially reduced, while the benefits of the physical presence of a passenger (such as clear transmission and access to non-verbal cues) were maintained, resulting in better performance overall.

Each of the articles reviewed above appears on the surface to provide strong evidence supporting the notion that passenger and cell phone conversations affect driving performance to a similar degree. However, closer analysis reveals important issues that provide a basis for arguing that the reverse may in fact be true. Research supporting this proposition will now be examined.

Research finding differences between passenger and cell phone conversations

Given the impact methodological differences may have had on the findings mentioned above, it may be helpful to begin the analysis of these studies by exploring the epidemiological research. McEvoy et al. (2005) conducted a large-scale case-crossover study in which the likelihood of being involved in a crash requiring hospitalisation when conversing on a cellular phone while driving was calculated. A second study was conducted simultaneously and examined the likelihood of being in a crash requiring hospitalisation when driving with passengers (McEvoy et al., 2007). Cases for each study included all consenting drivers who presented with an injury at one of three main hospitals in the area covered by the study. In the mobile phone study, cases acted as their own controls through the use of their telephone records. In the passenger study, four controls for each case were recruited from petrol stations in the vicinity of the crash in question, and were matched for time and day of the week. The first study revealed that cell phone use whilst driving was associated with a four-fold increased risk of a crash serious enough to require hospitalisation. The second study found that driving with passengers also increased the likelihood of a serious crash, but not to the same extent as mobile phone use. It was revealed that driving with two or more passengers doubled the risk of a serious crash compared to driving alone.

One weakness of McEvoy et al.'s (2007) passenger study was that the recruitment method used to obtain control cases may have resulted in a selection bias. It is possible that of the drivers approached at petrol stations, those with passengers may have been less likely to participate so as to not keep their passengers waiting unnecessarily. However, 17% of McEvoy et al.'s final control sample was carrying passengers, a figure equivalent to the national average for Australia, the country in which the study was conducted.

As mentioned earlier, a handful of laboratory studies have confirmed this finding that cell phone and passenger conversations have differing effects on driving performance and associated crash risk. However, before comparing these epidemiological findings to laboratory-based experiments, it must be taken into consideration that these two studies focused purely on associated crash risk rather than on differentiating between the potential factors underlying these associations. That is, it cannot be determined from these results alone what aspect of passenger carriage contributed most to the increased crash risk. Passengers can distract drivers in several ways, such as through movement, or misbehaviour in the case of children. They could also possibly contribute to a crash by accidently blocking the driver's view at a critical moment. Therefore, it is likely that these results are actually an underestimation of the differences between the effects of passenger and cell phone *conversation* on crash risk, and that if physical distraction by passengers could be controlled for, an even greater difference would be evident.

Drews et al. (2008) looked closely at conversational behaviour displayed by participants under different conversational conditions and presented convincing evidence that conversational form and content may be a key factor differentiating passenger and cell phone conversations. Forty-eight drivers drove a simulated version of a 26km highway with multiple on and off ramps. There were two lanes of traffic in each direction, and the surrounding cars changed speed and lanes frequently. Drivers were instructed to leave the highway after a particular rest area. Measures of driving performance included lane-keeping, speed, following distance and navigation task accuracy. Each driver completed half of the drive without conversation (single task) and the other half talking to a partner who either posed as a passenger or called on a cell phone (dual task). Driver and converser roles were randomly assigned, and the order of the single and dual task segments was counterbalanced. Pairs were instructed to share a close-call story that the other party had not heard before. With this as a catalyst, conversation soon became naturalistic and free-flowing. Conversation measures included the number of references to traffic; which party initiated the reference; and how many consecutive traffic-related turn takes this initial reference triggered.

Drivers in the cell phone condition exhibited poorer lane-keeping and a greater following distance, yet no change in speed across conditions was observed. Drivers in the cell phone condition were also four times more likely to miss the exit than those in the passenger condition. In regards to conversation content, it was found that fewer references to traffic were made in the cell phone condition, but further analysis revealed that the number of references initiated by the driver did not change, but rather conversers in the cell phone condition made significantly fewer. Analysis of turn taking showed that twice as many related turns were taken following a traffic reference in the passenger condition than in the cell phone condition.

Further explanations for the differences in driving performance noted between passenger and cell phone conversations were found through analysis of conversation form. Drews et al. (2008) found that while drivers and passengers didn't modulate the overall speed of their conversation in response to greater driving difficulty, they did alter the complexity of the conversation (as measured in syllables per word). Traditionally, conversation complexity has been incorporated as an independent variable as opposed to a dependent variable (e.g., Nunes & Recarte, 2002). This new approach suggests that changes in conversation complexity are a by-product of other causes of distraction rather than the key contributor. It is also possible that this measure may have been affected by a greater occurrence of short exclamations in the passenger condition. Given that fewer references to traffic were made in the cell phone condition, it could be that if the complexity of sentences not involving traffic references were analysed independently, less of a difference would have been observed.

While there was no noticeable overall change in speech rate as driving difficulty increased, an interesting pattern was observed when drivers' and conversers' speech production rates were analysed separately. As expected, drivers in the passenger condition decreased their production rate as driving difficulty increased, but drivers in the cell phone condition actually increased their production rate as driving demand shifted from low to moderate. This shows the importance of using naturalistic conversation between two participants as opposed to conversing with an experimenter. Managing turn-taking is an important task of conversation and tasks which eliminate the need for this (e.g., a driver responding to an experimenter's questions) do not accurately recreate the cognitive demands placed on a driver engaging in a real-life conversation.

These findings were supported by Charlton's (2009) simulator-based research. Sixty-four participants drove a simulated stretch of mainly rural road that included a series of hazards. At each hazard, speed, reaction time, time-to-collision (TTC), and crash rate data were collected as measures of driving performance. Performance was compared across four different conversation conditions: a no-conversation control; passenger; cell phone; and a remote passenger condition in which the conversation partner conversed with the driver via a cell phone but was still able to see the driving scenario as it unfolded in real-time. Participants were recruited in pairs so in each case the converser was known to the driver. Pairs were free to discuss whatever they wished, and the converser was supplied with a list of conversation topic ideas to use if needed. To determine the presence of conversation modulation, 20 second segments of the conversation surrounding the hazards were analysed for utterance length, number of pauses longer than two seconds, and percentage of situation awareness utterances (utterances within the given time period pertaining to the immediate driving situation). Driver and converser scores were analysed independently.

Charlton (2009) found that in general, those in the control and passenger groups displayed significantly slower speeds at hazards than those in the remote passenger and cell phone groups. At some hazards, those in the passenger group actually drove slower than those in the control group. Cell phone and remote passenger conversation resulted in significantly longer reaction times and shorter TTCs when compared to passenger conversation and the control group. In fact, in several instances those in the cell phone group made no deceleration response to an upcoming hazard and this may have contributed to this group displaying the highest crash rates.

Charlton (2009) suggests conversational differences provide the explanation for the differing levels of driving performance described above. For all three discourse measures, significant differences between the means were observed between conversers in the cell phone and passenger groups, as well as the cell phone and remote passenger groups. Conversers in the passenger group were significantly different from the remote passenger group for all measures other than utterance length. In the case of drivers, the passenger and cell phone groups were significantly different from each other for both mean number of pauses and mean percent of situation awareness utterances, and approached significance for mean utterance length. Drivers in the remote passenger group were significantly different from both other groups for situation awareness utterances, but were comparable to the passenger group for number of pauses, yet similar to the cell phone group for utterance length.

Charlton's (2009) findings imply not only that passenger and cell phone conversations have significantly different effects on driving performance, but also that conversation differences are a key factor contributing to these differences. However, the inclusion of the additional remote passenger condition revealed results that suggest there are obviously other elements in play. Support for this idea can be found in one of the discourse measures, utterance length. For drivers, there was no significant difference in mean utterance lengths between the remote passenger and cell phone conditions, with drivers in each of these two conditions speaking in significantly longer utterances than drivers in the passenger group; whereas for conversers, there was no significant difference in utterance lengths between the passenger and remote passenger conditions, while both were significantly lower than the cell phone condition. That is, drivers spoke in shorter utterances when the converser was in the vehicle, whereas conversers spoke in shorter utterances when they were able to see the road. This suggests that the pre-requisites for certain aspects of conversation modulation differ for drivers and conversers, an idea that was first proposed by Crundall et al. in 2005.

Crundall et al. (2005) investigated differences between passenger and cell phone conversations and their associated effects on driving performance in an on-road field test. Driving difficulty was manipulated by having participants drive a circuit that included suburban, urban, and rural roads, as well as dual carriageways. Conversation conditions included passenger, mobile phone and an additional condition unique to this experiment, blindfolded passenger. A conversation game was prescribed in which the passenger had to subtly get the driver to say certain words while the driver had to guess which words the passenger was trying to get them to say. The nature of the game allowed for relatively naturalistic conversation. The number of utterances, utterance length, and number of questions asked were used as conversation measures.

Results showed that as driving difficulty increased, suppression of conversation by both parties occurred in the normal passenger condition, but not in the mobile phone condition. This suppression was characterised by a reduction in all conversation measures. Interestingly, drivers also reduced their number of utterances during high-demand driving in the blind-folded passenger condition. Crundall et al. (2005) concluded that drivers' conversation is more affected by the physical presence of a passenger, whereas a partner's conversation is more affected by their ability to see the driving situation. Despite this being supported in part by Charlton's (2009) research, there are several reasons to treat this deduction with caution when considering this research on its own merits. The first is that instead of comparing all conversation conditions separately, the mean of the no-vision conditions (blind-folded passenger and mobile phone) was compared to the normal condition for the majority of the analyses. Yet both drivers and conversers produced a very different pattern of results in each of the two no-vision conditions, which suggests that Crundall et al.'s decision to combine these two means was an inappropriate choice for meaningful analysis of these results. In addition, the judgement regarding conversation partners was based on marginal interaction effects between road type and conversation condition but overall, the partners' conversation patterns matched those of the drivers closely. This suggests that factors other than seeing the road play an important role for the converser as well. It could be that psychological closeness is a necessity, or perhaps that modulation is a safety-motivated, so that in times when the converser's safety is not at risk (i.e. during a cell phone conversation), they are less inclined to alter their conversation.

Another reason to interpret these results with caution is that driving performance was not measured so it is unknown whether this suppression actually altered driving behaviour in any way. Also of concern is that only twenty participants were recruited for the study. These participants were then placed into pairs, limiting the sets of data to just ten. To further complicate matters, half of the participants' data was affected by interference between the video equipment and the mobile phone, so analysis was carried out on just five data sets. This is an obvious weakness of the study, and it may be that the significant results found in this study were the result of large differences between one or two participants. However, it may also be possible that the effects described by the authors do in fact exist, and had the study included more participants, the differences between driver and passenger conversation patterns, especially regarding the blindfolded passenger condition, may have been much more noticeable. It is therefore still prudent to consider these results and their implications when examining passenger and cell phone conversation research as a whole.

Lastly, Hunton and Rose (2005) also investigated differences between passenger and cell phone conversations but proposed that cell phone and passenger conversations differ in more ways than purely discourse-related measures. They hypothesised that the lack of non-verbal cues available during a cell phone conversation contributes to the distraction associated with cell phone use while driving, as drivers must devote more cognitive resources to the conversation to compensate for this lack. They also suggested that communication training can result in improved driving performance while conversing on a cell phone, regardless of the presence or absence of conversation modulation. They tested these hypotheses by comparing driving performance of participants with and without basic pilot training on a simulated track across three different conversation conditions: no conversation, passenger conversation, and hands-free cell phone conversation. Pilots were used as a fundamental element of pilot training is learning to develop effective radio communication skills.

During the experiment, conversation modulation was controlled for by using trained experimenters as the conversation partners in all trials. The experimenters used scripts and were trained to keep the driver constantly engaged in conversation, thus ensuring all participants were exposed to the same conversational conditions. The experimenters were not allowed to look at the simulation to avoid any chance of unintentional conversation modulation.

Driving performance was assessed by the number of incidents (e.g., speeding, failing to give way, following too close) and the number of crashes.

The results showed that pilots produced better driving performance than their non-pilot counterparts during the cell phone condition, despite both groups having similar performance levels in the no-conversation control condition. To explore possible reasons for this difference, participants were asked to provide ratings on several factors relating to non-verbal aspects of a conversation. These self-ratings revealed that during cell phone conversation, non-pilots had a significantly higher desire to see the conversation partner than did pilots, and non-pilots reported utilising a significantly greater quantity of cognitive resources trying to visualise the conversation partner's face. It was also discovered that non-pilots experienced a significant increase in anxiety levels between the no-conversation condition and passenger conversation, and a further increase in anxiety between passenger and cell phone conversation conditions. Pilots showed no significant increase in anxiety levels between the no-conversation and passenger conditions, but did show a significant increase in anxiety from these two conditions to the cell phone condition. While both pilots and non-pilots experienced an increase in anxiety when engaging in a cell phone conversation while driving, the extent of this increase was significantly higher for the non-pilots than pilots.

Considering conversation modulation was controlled for in this experiment, these results show that the presence or absence of conversation modulation cannot be solely responsible for the decrease in driving performance associated with cell phone use while driving. Hunton & Rose's (2005) findings suggest non-verbal cues play an important role in alleviating distraction associated with cell phone use, and support their hypothesis that at least some of the associated distraction can be attributed to the driver using valuable cognitive resources to produce a mental image of the converser's face. However, it is important to note that these particular findings are based on self-reports only. To truly examine the effect of non-verbal cues and mental imagery in driver distraction during remote conversations, it is necessary to design an experiment that empirically compares driving performance under two different remote conversation conditions – one during which the driver has access to non-verbal cues and one in which they do not. That is the aim of the present research.

The current study – outline, research questions, and hypotheses

A remote conversation condition lacking non-verbal cues is relatively easy to create – a normal cell phone conversation meets this criterion. A remote conversation condition incorporating non-verbal cues is slightly more difficult, but can be found in the form of videophone conversation. A two-way videophone allows both driver and converser to see the other person's face throughout the conversation, thus eliminating the need for the driver to produce a mental image of the converser's face and surroundings. If Hunton & Rose's (2005) suggestions are correct, a videophone conversation should result in improved driving performance when compared with a cell phone conversation. This also ties back into Gugerty et al.'s (2004) claims that seeing the driver's face is crucial.

There are several important factors that must be considered when incorporating a videophone into a driving scenario. One possibility that cannot be ignored is that introducing an in-vehicle display may in fact increase distraction rather than decrease it, especially if the driver feels obliged to attend to the display. However, a study by Tsimhoni and Green (2001) into driver's glance patterns when using display-intensive in-vehicle navigational devices revealed that as the visual demand of the road increased, drivers' glances at the display became shorter and less frequent. Thus it could be hypothesised that should the driver be attending to the road at least in part, they will modify their glance patterns when they encounter hazards and thus minimise the distracting effects of the screen.

This in turn could actually result in a further potential benefit of using a videophone. Conversation modulation is usually defined as the parties involved altering the pace and form of a conversation in response to the demands of the driving situation. Yet it is also possible that even when they can't see the road, a conversation partner may alter the pace and form of a conversation in response to an indirect trigger, such as a reduced glance rate or other non-verbal cues displayed by the driver as they react to real-time driving demands. As such, rather than controlling for conversation modulation as Hunton and Rose (2005) did, this study will utilise naturalistic conversation, and retrospectively analyse conversation for evidence of modulation.

In summary, this study focussed on two main areas of investigation: the effects of concurrent video phone use on driving and the possible reasons for these effects. To address these areas, two specific research questions were developed, which were: (1) does conversing via video phone whilst driving result in improved driving performance when compared to driving while conversing on a hands-free cell phone?; and (2) does the use of a videophone result in conversation modulation by a remote converser? Due to the lack of research investigating the role of non-verbal cues and mental imagery in distraction associated with cell phone use while driving, it is difficult to state hypotheses that favour a particular outcome. Rather, in this case it is more appropriate to outline possible outcomes, of which there are three.

1. The use of a videophone will result in worse driving performance, indicating that the in-vehicle display adds a further source of distraction;
2. The use of a videophone will produce driving performance comparable to that produced by cell phone conversations, indicating that either no benefits are produced by the videophone or that any benefits are nullified by simultaneous added distraction; or
3. The videophone will result in improved driving performance compared to a cell phone conversation, indicating that access to non-verbal cues can alleviate some of the distraction associated with cell phone use. Should this be the case, further analysis will be used to determine the way in which non-verbal cues and/or conversation modulation contributed to this improvement.

Method – Experiment 1

Participants

Participants were recruited in same-gender pairs with a friend or acquaintance to ensure fluent, comfortable conversation in the conversation trials. A total of 29 pairs were recruited for the study, but due to technical issues with the simulator and the occurrence of simulator sickness in some participants, only 21 pairs were used in the final analysis. Of these pairs, 7 were male and 14 were female. The final sample had a mean age of 29.19 years, an average of 13 years driving experience, and drove an average of 151 kilometres per week. All participants were fluent in English, with New Zealand Europeans making up 88% of the sample, and the remainder including New Zealand Maori, English, Samoan, and Asian participants. All participants held a full NZ drivers’ licence, and were owners of a cell phone. The sample included University of Waikato undergraduate psychology students and members of the general public. University of Waikato students were recruited via notices on university notice boards (both physical and electronic), as well as announcements in tutorials. Members of the general public were recruited via notices placed on community notice boards and through social networking sites. Undergraduate psychology students were compensated for their time through the receipt of course credits. Members of the general public received their choice of a coffee or grocery voucher of $4-$5 value.

Apparatus

The experiment was conducted in the University of Waikato driving simulator. The simulator consisted of a BMW 316i, with its engine removed. The simulated roads were projected onto three angled surfaces – a central one positioned 2.42m directly in front of the driver's seat, which was flanked by two peripheral surfaces, each connected to the central surface at angles of 62°. Each of these surfaces was angled away from the driver at 14° from bottom to top. Three corresponding projectors cast an image on each surface, with the central image measuring 2.64m wide and 2.10m high and each peripheral image measuring approximately 2.65m by 2.00m. This set-up created a horizontal visual angle of 175° and a vertical visual angle of 41°. (Participants were able to adjust the seat slightly to suit therefore this is an approximation based on the above distances and angles.) At 1900 by 1200 pixels, the central image had a higher resolution than the peripheral images, each measuring 1024 by 768 pixels. To create a full 360° simulated effect, rear view and driver's wing mirror images were displayed using two colour LCD screens mounted accordingly, each measuring 12.065cm by 7.493cm and having a resolution of 640 by 480 pixels. The speedometer and revolution counter located on the vehicle's dashboard were configured to provide accurate real-time information to the driver, and the power-assisted steering wheel was equipped with force-feedback to emulate the resistance felt in an actual driving situation. Realistic engine and road noise was broadcast via four in-vehicle speakers and a subwoofer located underneath the vehicle. In addition to speed and lane position, the simulator's software recorded all driver control actions. This information was collected continuously throughout the drive.

The simulator was fitted with an internal hands-free phone connection which was used for the Cell phone and Videophone conditions. A digital camera was mounted in the rear of the vehicle and was used to capture all experimental trials for analysis and review purposes. A second digital camera was mounted on the centre console of the simulator for the Videophone condition, with an additional LCD screen mounted adjacent to this camera. An additional projector was used to cast a moving image on the car windows behind the driver's head, to ensure it appeared to Videophone conversers that the driver was indeed driving. Care was taken to angle the image so that it appeared in the converser's field of view (via video camera) but did not impede or distract the driver in any way. The image cast was a DVD replay of the image displayed on the right-hand screen of the simulator during a test drive of the experimental road and as such did not always match the participants' speed or location exactly. External and internal views of the University of Waikato driving simulator are shown in Figures 1 and 2 on the following page.

The simulated driving scenario

The simulated section of road used in this experiment was an accurate geometric representation of a 23.2km section of New Zealand's State Highway 2, initially constructed for Charlton's 2009 driving experiment. At the time of its construction, care was taken to ensure all road signs, roadside objects, and general road engineering characteristics were incorporated into the simulation accurately, based on information obtained from road controlling authorities and

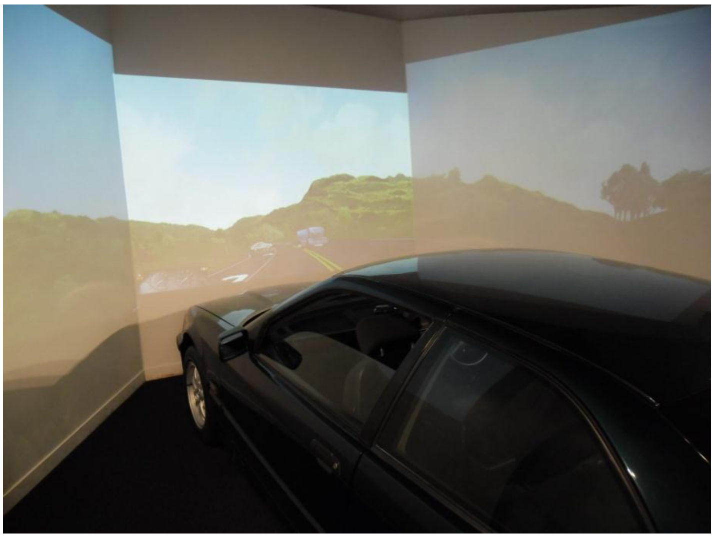


Figure 1. The University of Waikato driving simulator.

Figure 2. Internal view of the University of Waikato driving simulator, showing set up for Videophone condition.

from video recordings of the actual highway, filmed from the driver's perspective. As part of his aforementioned experiment, Charlton also incorporated five fictitious traffic hazards, typical of what may be encountered on a rural stretch of New Zealand road. Each of these hazards was retained for this experiment, with a few slight alterations. The simulation scenario for the current experiment is described below.

Participants began their drive in a 100km/h speed zone, indicated by a sign located on the left hand side of the road, 100m in front of their starting position. Other traffic, including cars, light trucks and heavy vehicles, was present throughout the entirety of the drive, at a rate corresponding to 8000-10000 passenger car units per day. The first hazard was located 2.48km into the drive and involved a busy "T" intersection where a side road met the main highway from the left. A large truck was waiting to turn across the participant's path into the side road, and a police car was waiting to turn from the side road onto the main highway, also across the participant's path. The police car was stopped slightly forward of the give way lines, creating what appeared to be a tight gap between the truck and police car through which the participant had to navigate. Prior to reaching this intersection, the participant encountered two heading vehicles, each travelling at 95km/h. Approximately 150m before the intersection, the car immediately in front signalled a left turn and moved into the left turning bay. The remaining leading car's brake lights were illuminated momentarily before it proceeded through the intersection. After passing through the intersection this car increased its speed to 135km/h then pulled over 2km later and was removed from the driving scenario.

The second hazard occurred 4.5km into the drive, and involved a hidden car suddenly pulling out onto the highway in front of the participant. The event took place in a 70km/h speed zone, which was indicated by speed limit signs placed on either side of the road approximately 200m before the hazard. The entering car was parked outside a pub and was obscured from the participant's vision by a second parked vehicle until moments before it entered the highway.

The hazard was designed so that if the participant slowed to 70km/h as per the posted limit, the parked car could enter the road safely without the need for the participant to brake. However, if the participant did not slow down, evasive action (either braking or swerving) was required to avoid a collision. Two hundred metres later the speed limit increased back to 100km/h.

A third hazard appeared just over 4km further down the road, in the form of a one lane bridge. Approximately 550m before the bridge, a 60km/h speed limit was introduced, indicated by signs on either side of the road. Warning signs posted 272.5m and 42.5m prior to the bridge alerted the driver to its presence and indicated the participant had right-of-way. However, as the participant approached the bridge a large truck travelling in the opposite direction exited the bridge, followed by two cars. A third car came to a rolling stop on the other side to allow the participant's vehicle to cross. As per Hazard 2, if the participant slowed to the correct speed limit of 60km/h, they could navigate the bridge with little interference from other traffic; however if they maintained a higher speed, they were forced to stop and wait for the other traffic already on the bridge. The speed limit increased back to 100km/h 250m after the bridge.

The next major feature the participant encountered was an overtaking lane, 3.15km later. Although not specifically considered a hazard, the presence of other vehicles in the overtaking lane presented a challenging driving situation for most participants. Warning signs were located 2km, 1km, and 400m prior to the start of the lane. Approximately 1km before the participant reached the overtaking lane, they encountered a car and van on the road in front, both travelling at 87km/h. These vehicles soon caught up to a tanker travelling at

78km/h. The road layout and strategic positioning of oncoming traffic prevented the driver from being able to overtake the procession before reaching the overtaking lane. When the lane was reached, all three leading vehicles moved left initially, before the car and van quickly signalled right and moved into the overtaking lane to pass the tanker before the participant was able to. The van pulled left after passing the tanker, while the car stayed in the right lane until it had passed both the tanker and the van. Given the speeds of the leading vehicles during this manoeuvre (96km/h) and the length of the overtaking lane (880m), the participant had time to safely overtake the van and tanker but not the other car before the overtaking lane ended.

Hazard 4 consisted of road works, and was located 5.28km beyond the end of the overtaking lane, on the crest of a hill. A temporary 30km/h speed limit sign was located 60m prior to the start of the road works, followed by a loose gravel sign 30m later. A series of road cones placed on the median line marked the 123.5 metre-long road working site, the first 38.5m of which consisted of loose gravel, followed by 85m of sealed road with no road markings. A “works end; 100km/h” sign marked the end of the hazard.

The final hazard involved a slip that extended 1.5m into the participant’s lane, and was situated 3.5km beyond the road works. A “slip” warning sign was present 100m prior to the slip’s location but this hazard was not accompanied by a reduction in speed limit. Cones redirected the participant around the slip.

The final feature of the road appeared 1.48km later in the form of a rest area on the left hand side of the road, complete with picnic tables and rubbish bins. The rest area was preceded by a sign alerting the driver to its presence

300m before it began, and was also marked with a sign at its entrance. Participants were told at the start of their drive to pull into this rest area when they came to it, and that this would mark the end of the experimental drive. Figure 3 shows a map of the simulated driving scenario, including speed limits, hazard placement and the rest area. Images of the five hazards and rest area are shown in Figures 4-9.

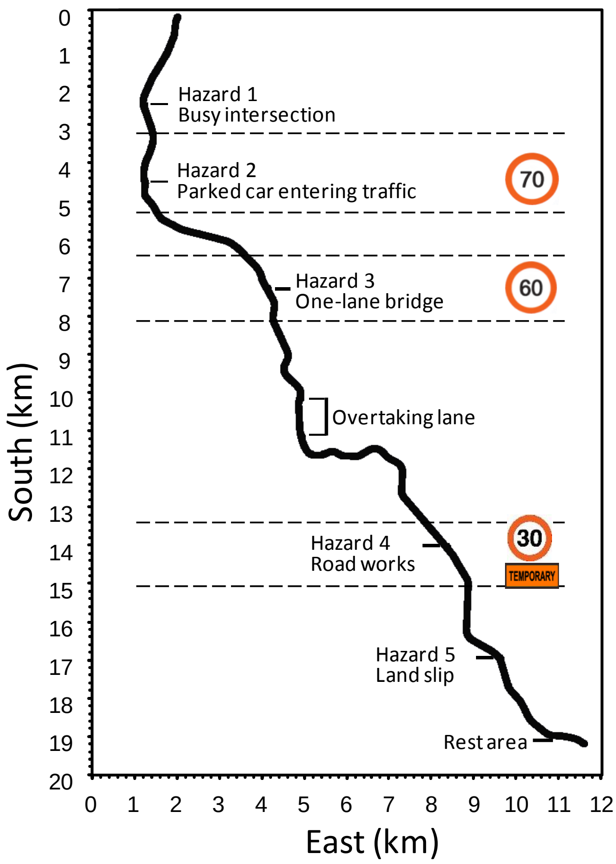


Figure 3. Map of Experiment 1 simulation scenario, showing speed limits and placement of hazards and rest area. All areas with unspecified speed limits are 100km/h zones.

Figures 4 and 5. Hazard 1 – busy intersection with turning police car (left panel) and Hazard 2 – previously obscured parked car pulling onto the road (right panel).

Figures 6 and 7. Hazard 3 – one lane bridge (left panel) and Hazard 4 – road works (right panel).

Figures 8 and 9. Hazard 5 – slip (left panel) and the rest area (right panel).

Procedure

The experiment consisted of four conditions: a no-conversation control condition, an in-car passenger conversation condition, a hands-free cell phone condition, and a videophone conversation condition. Participant pairs were randomly assigned to a condition at the time of recruitment, with six pairs being assigned to each of the conversation conditions, and three pairs to the control condition. For the conversation conditions, one member of the pair acted as the driver and one as the conversation partner. For the Control condition, both members of the pair acted as drivers, thus only three pairs were required. The session began with a brief explanation of the aims of the research and what was involved. Following this, one participant completed a short practice drive while the other filled out an informed consent form and a brief questionnaire about their driving and talking habits. Participants subsequently swapped places, allowing each participant to experience a brief test drive. Participants then self-selected who would be the driver for the experiment and this person was given further practice as described below. For the Control condition, each participant completed the experiment individually while the other member of their pair waited in a waiting room. As such, the initial test drive to decide the driver was omitted.

The same section of road as that used in the experimental drive explained above was used as the basis for the practice road, which allowed the experimenter to accurately gauge when the participant had completed enough practice to successfully negotiate the experimental drive. The scenery and landmarks were altered significantly enough to ensure the experimental road did not seem familiar to the participant. A second practice road, consisting of a straight road with a series of road cones placed 100m apart on the centre line, was available to give participants additional steering-specific practice. Participants were asked to weave through the cones as smoothly as possible at a range of different speeds. Practice continued until both the experimenter and the participant felt the participant was ready to begin the experiment (usually about 10-20 minutes). The experiment involved participants driving the experimental road whilst engaging in conversation in whichever format was required for their randomly-assigned condition. For the Passenger condition, the conversation partner was seated in the simulator next to the driver and was able to see the road and its surroundings. In the Cell phone condition, the conversation partner was seated in an adjacent room, unable to see the road or the driver. For the Videophone condition, the conversation partner was seated in the adjacent room in front of a computer screen displaying video feed of the driver's face, as captured by the camera mounted on the dashboard. A small camera was located in front of the computer screen and sent a live image of the conversation partner through to the LCD screen mounted on the dashboard. As per the Cell phone condition, the conversation partner in the Videophone condition was also unable to see the road.

Participants were asked to start their conversation by completing one of three conversational tasks – composing a list of 12 items they agree to take to a deserted island; identifying 12 songs they would both agree to include in a playlist for a road trip; or creating a grocery list for seven home-cooked dinners they would both be happy to eat over the next week, while staying within a budget of $120. These tasks were designed to get the conversation flowing initially and give participants a topic to return to should they run out of conversation ideas later in the drive. Participants were informed that should their conversation flow on to other topics naturally while carrying out the above task, they were welcome to move on without completing the task. Participants were required to begin their conversation at least two minutes before the driver began driving so that non-driving conversation baseline could be established.

At the end of their session, drivers were asked to provide a difficulty rating for both the drive and the conversation based on Charlton's seven-point mental workload/driving difficulty scale (2004, cited in Charlton, 2009). They were also asked to give a rating of interference caused by the conversation. This was also based on a seven-point scale, with 1 representing no interference and 7 representing complete interference to the point they could no longer drive. Conversation partners were also asked to give a difficulty rating for the conversation from their perspective. In addition, they were asked to rate how much they felt the conversation interfered with their partner's driving and to describe the factors they used to determine this level of interference. Ratings of driving difficulty only were collected from those in the Control condition.

Data collection and analysis

Each of the five hazards the drivers encountered served as a data collection point. Speed was recorded at each of the hazards, as were reaction times and times-to-collision. Reaction times were measured from a specifically chosen Zero Reaction Time (RT = 0) location and a reaction referred to the first time a driver removed their foot from the accelerator prior to passing the hazard.

Times-to-collision were calculated to determine the appropriateness of the above reactions. The number of crashes was also recorded. The parking area was used as a gauge of strategic driving performance, and whether or not participants entered the area was noted. The conversation was recorded and 20 second segments, beginning 10 seconds before each hazard, were transcribed for analysis. Discourse measures included the mean utterance length (measured in number of words), the number of pauses longer than 2 seconds, the number of words spoken by dyads during the 20 seconds and the number of utterances pertaining to the hazard or road (situation awareness utterances). One-way multivariate analyses of variance were used to identify any main effects of group in regards to speed at each of the five hazard points. These were followed by univariate analyses and post hoc pair-wise comparisons to clarify any differences or interactions. The same procedure was employed to analyse reaction time and times-to-collision, as well as the majority of conversation measures. Some measures, such as situation awareness utterances and ratings, required only univariate as opposed to multivariate analysis. These one-way analyses of variance were also followed by post hoc comparisons to clarify any differences or interactions. Partial eta-squared (η_p^2) was used as a measure of effect size, with a η_p^2 value of 0.01 representing a small effect size, 0.06 a medium effect, and anything greater than 0.14 being considered a large effect (Cohen, 1988).

Results – Experiment 1

Driving Performance

Speed. The mean speeds of drivers in each of the conversation conditions were compared at a series of successive location points surrounding each of the five hazards, as displayed in the graphs in Figures 10-14. Four drivers failed to complete the experiment due to either technical difficulties with the simulator or because they were unable to continue following a crash, resulting in missing data at the later hazards. In all cases, missing data was replaced with the group mean.

As can be seen in the graphs, drivers in the Cell phone condition displayed the slowest speeds at all hazards except Hazard 3. No clear pattern was evident in the speeds of the remaining three conversation conditions, with the Videophone condition producing the fastest speeds at Hazard 1; the Control condition producing the fastest speeds at Hazard 2 and 5; and the Passenger condition producing the fastest approach speeds at Hazard 3. The mean speeds of those in the Videophone and Passenger conditions were equally fast at Hazard 4. Error bars showing 95% confidence intervals were not included in these graphs as differing levels of within-group variation meant the conversation condition displaying the highest mean speed at a given point did not necessarily have the highest confidence interval value. As such, including them would have hindered rather than aided understanding of the data.

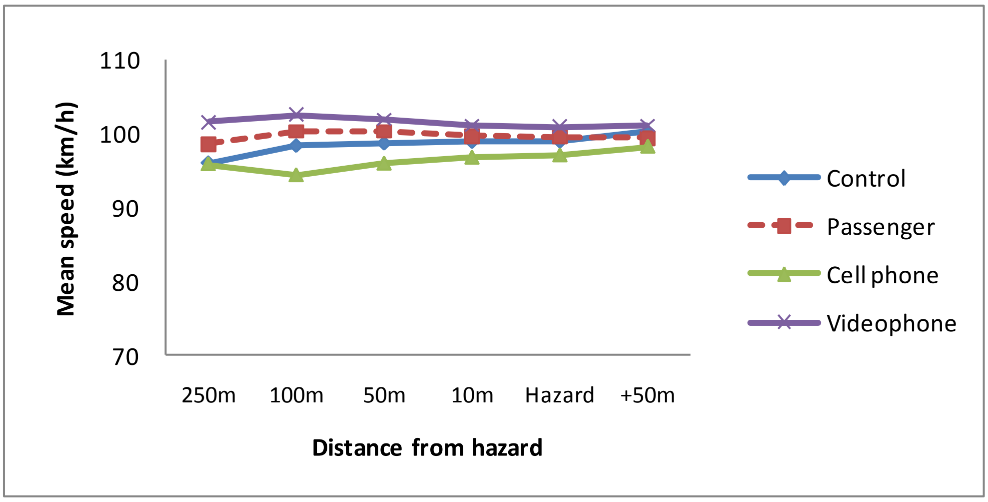


Figure 10. Mean speeds at Hazard 1 - busy intersection.

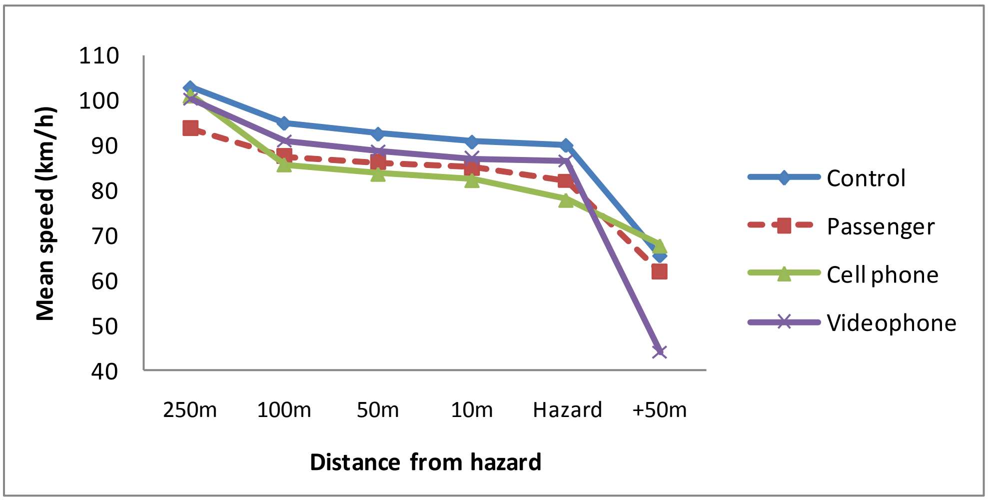


Figure 11. Mean speeds at Hazard 2 - parked car entering traffic.

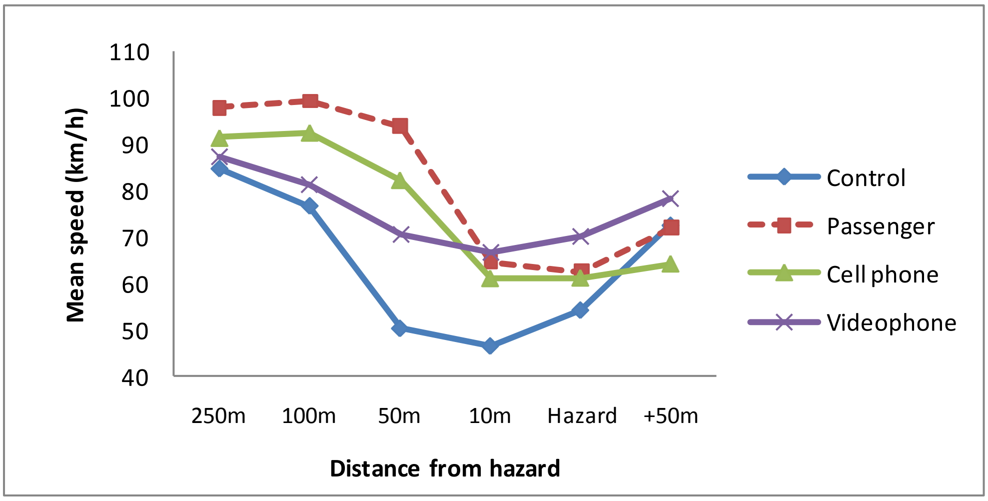


Figure 12. Mean speeds at Hazard 3 – one lane bridge.

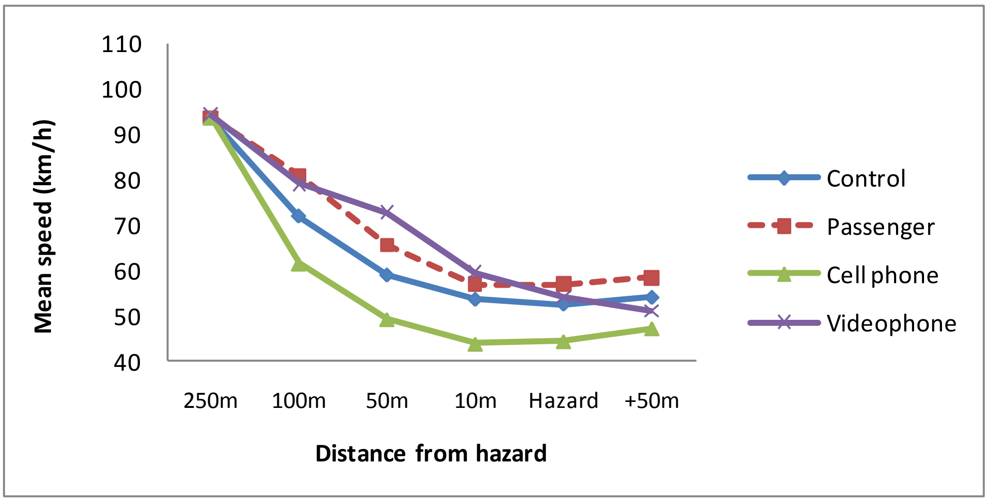


Figure 13. Mean speeds at Hazard 4 – road works.

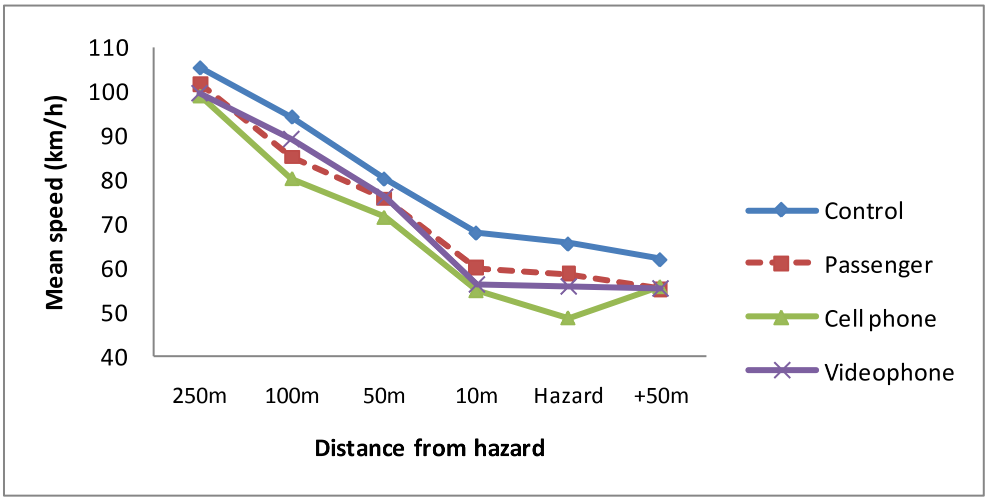


Figure 14. Mean speeds at Hazard 5 – slip.

The participants' mean speeds at each of the hazard locations were compared using a 4 x 5 mixed design multivariate analysis of variance with four levels of the between-subject factor, conversation condition, and five levels of the within-subject factor, hazard location (General Linear Model, PASW Statistics 18). The analysis failed to reveal any significant differences between the conversation conditions ($F_{(3,20)}$ = 0.707, p > 0.05, η_p^2 = 0.096), or any significant interaction between conversation condition and hazard ($F_{(7.7, 51.335)}$ = 0.539, p > 0.05, η_p^2 = 0.075, using the Greenhouse-Geisser adjustment for sphericity).

A series of additional one-way multivariate analyses of variance of identical design were used to compare the participants' speeds at each of the locations measured at each hazard individually (due to the small sample size), and these also failed to reveal any significant differences between any of the conversation conditions 50m beyond the hazard ($F_{(3,20)}$ = 0.188, p > 0.05, η_p^2 = 0.027), or at any of the points measured leading up to the hazard (ps > 0.05). There was, however, an interaction between hazard and conversation condition at one of the approach points, and a subsequent one-way analysis of variance with post hoc pair-wise comparisons made using the Bonferroni adjustment method showed the mean speeds of drivers in both the Passenger and Cell phone conditions were significantly higher than those of drivers in the Control condition 50m prior to Hazard 3 (p < 0.01 and < 0.05 respectively).

The reason for significant differences being present at this particular location but no others becomes clear when the nature of the hazard is taken into account. Hazard 3 involved a one-lane bridge, located within a 60km/h zone. The hazard was designed in such a way that if participants slowed to the speed limit, they would arrive at the bridge just as a line of traffic travelling in the opposite direction finished exiting the bridge. This allowed the participant to enter the bridge without stopping, and participants who exhibited this response thus maintained an approximate speed of 60km/h as they approached and passed through the hazard. However, many participants failed to recognise or respond to the change in speed limit and maintained an approximate speed of

100km/h as they approached the hazard. These drivers were then forced to come to a complete stop at the bridge to allow the traffic travelling in the opposite direction to exit, meaning they exhibited very high speeds further away from the hazard and very low speeds at points closer to the hazard. Some drivers also failed to notice that the bridge was in fact one lane and not two, and entered the bridge at speeds of close to 100km/h, colliding with the traffic travelling in the opposite direction. These three possible responses resulted in within-group variation being as great as 90km/h at the hazard site for some of the conversation conditions. Also, given the low numbers of participants in each conversation condition, one or two participants exhibiting an extreme response at either end of the spectrum in one particular condition could result in this conversation condition being significantly different from another. These results highlight some obvious flaws in the experimental design which will be discussed in more detail in later sections.

Reaction time and time-to-collision. Reaction times were measured at each of the five hazards. A Zero Reaction Time (RT=0) location was chosen for each individual hazard, and a participant's reaction time was the number of seconds that elapsed from when the participant passed this RT=0 point until the moment they first took their foot off the accelerator. The RT=0 locations were carefully chosen with respect to surrounding road features so as to minimise the chance of including reactions that were in fact a response to other road features or excluding reactions that were actually in response to the hazard. For Hazards 1 and 5, RT=0 was the point at which the hazard first came into view. This was 220m prior to Hazard 1 and 425m prior to Hazard 5. Hazard 2 was situated 200m after the start of a new speed zone so in order to differentiate between reactions to the speed limit change and reactions to the hazard, RT=0 for Hazard 2 was located 50m after the 70km/h sign, 150m prior to the hazard. For Hazard 3, RT=0 was in line with the first warning sign for the one-lane bridge, located 280m before the bridge itself. For Hazard 4, RT=0 was located 300m prior to the road works.

The time-to-collision measure was used as a means of determining the appropriateness of a given reaction. It involved calculating the time it would have taken a participant to reach the hazard from the location at which they reacted had they maintained the speed they were doing at the exact moment they made the reaction. Thus, two participants with a reaction time of five seconds may have very different times-to-collision if one was travelling at 60km/h and the other 100km/h.

Figure 15 shows the mean reaction times and times-to-collision for groups at Hazards 2-5. Only three of the 24 drivers registered a response as they approached Hazard 1 so this hazard was excluded from the statistical analysis. Response rate was also low for Hazard 2, however initial analysis suggested participants were responding extremely late to this particular hazard, and as such, any deceleration responses occurring within the 50 metres immediately following the hazard were also included (hence the negative times-to-collision for this hazard). The graphs suggest conversation condition means were similar for both reaction time and time-to-collision at most hazards, with the exception of Hazard 3. The relatively large error bars suggest high levels of within-group

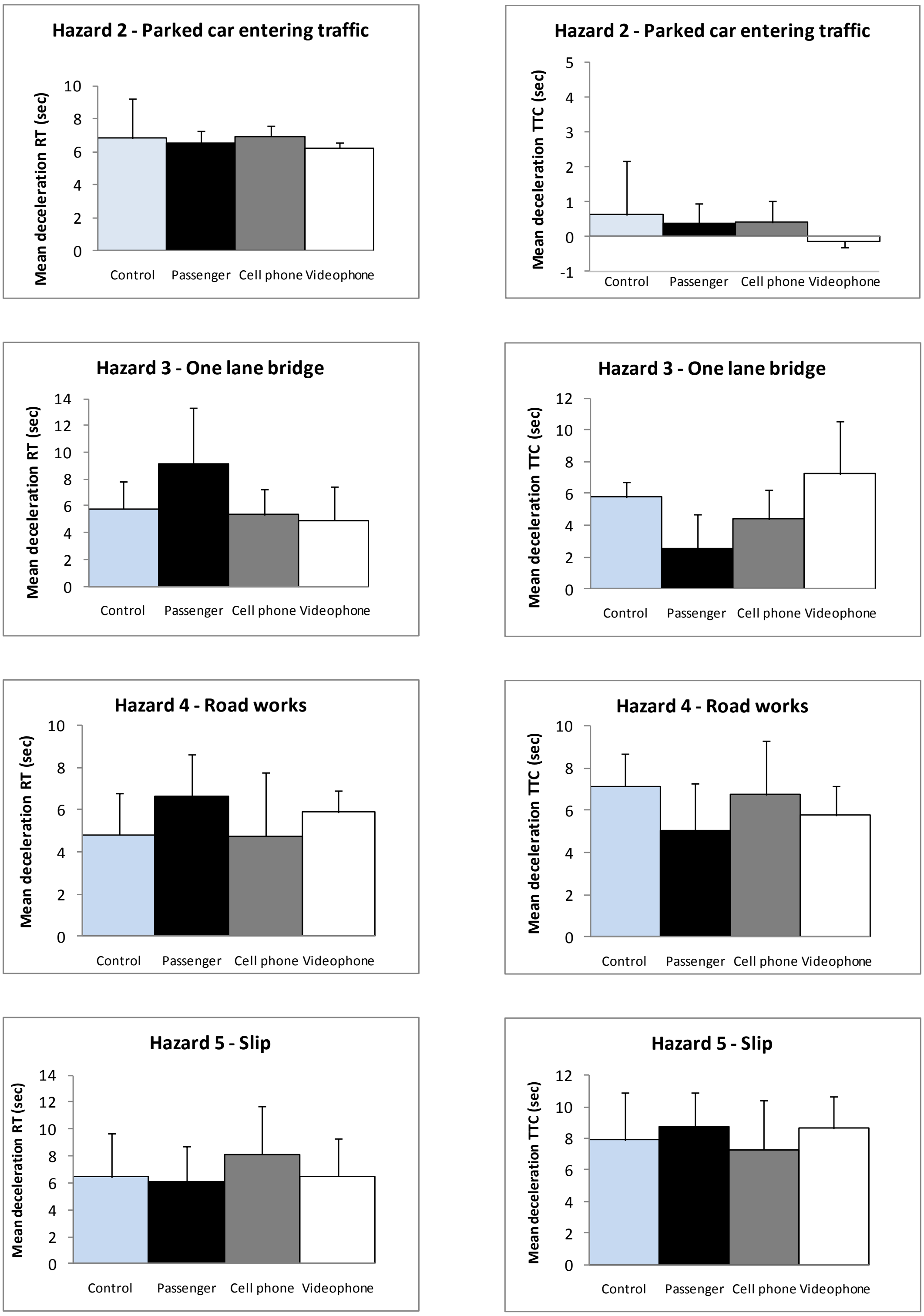


Figure 15. Mean reaction times and times-to-collision for Hazards 2 – 5 by conversation condition. Error bars show 95% confidence intervals. Note the negative time-to-collision values for the Videophone condition at Hazard 2 are a product of hazard boundary definition as explained in the text.

variability. This problem was present in a large number of measures in Experiment 1, and is addressed in more detail in the next section.

A mixed design multivariate analysis of variance of the conversation conditions' mean reaction times at Hazards 2-5 failed to show any significant differences between condition means for reaction time at any of the hazards ($F_{(3,20)} = 0.960$, $p > 0.05$, $\eta_p^2 = 0.126$). This analysis also failed to reveal any significant interaction between conversation condition and hazard location for reaction time ($F_{(9,60)} = 1.822$, $p > 0.05$, $\eta_p^2 = 0.215$). Due to the small sample size, a mixed design multivariate analysis of variance for times-to-collision was conducted separately and this also failed to show a statistically significant difference between the conversation conditions ($F_{(3,20)} = 1.810$, $p > 0.05$, $\eta_p^2 = 0.214$). A significant interaction between hazard and conversation condition was observed for this measure, and univariate analysis with Bonferroni-adjusted post hoc pair-wise comparisons showed that the mean time-to-collision of the Passenger condition was significantly lower than that of the Videophone condition at Hazard 3 ($p < 0.01$). No significant differences were found between any of the other conversation conditions at any of the other hazards (ps > 0.05).

There were several cases in which participants did not register a deceleration response, and these cases fell into two distinct categories. The first included participants who did not react because they had already removed their foot from the accelerator prior to crossing RT=0 and then maintained this response until the hazard. In this case, because of the carefully considered placement of the RT = 0 location points, it was assumed the participant had initially reacted to some other road feature and then at some point made a decision to continue this response in order to negotiate the upcoming hazard successfully. As it was impossible to determine exactly when this decision was made, these participants were assigned the mean reaction time and time-to-collision for their conversation condition.

The second category included participants who did not register a deceleration response between RT=0 and the hazard, but not as a result of reacting prior to RT = 0. In these cases, speed and video data was analysed to assess whether the participant was travelling at such a speed that they did not need to react, or whether it could be assumed that they did not react because they failed to see and respond correctly to the hazard. In the case that a participant's speed suggested they did not need to react, they were assigned the group mean for both reaction time and time-to-collision. In the case that a participant failed to react despite needing to in order to negotiate a hazard safely, this participant was assigned a maximum reaction time (based on the time taken to travel from the RT=0 location to the hazard if travelling at the speed limit) and a time-to-collision of zero.

Crashes. Despite only 24 driving participants being involved in this experiment, 25 crashes were recorded. Not all participants experienced a crash; however some participants had as many as three crashes. Most crashes occurred at Hazards 2 and 3 (the parked car entering traffic and the one-lane bridge), with a small number of participants losing control and crashing on a stretch of road between Hazards 3 and 4. A crash was defined as an event in which the participant's car made contact with another object, whether it was another vehicle, a road feature (e.g., a bridge), or part of the landscape (e.g., a participant losing control and colliding with a hill on the side of the road). Also included were any instances in which a participant's vehicle spun 180° or more as a result of them losing control. Instances where participants hit road cones were not considered to be crashes, partly because verification of this was difficult and also because an event of this type is usually negligible in real-world driving.

The high number of crashes recorded suggests a lack of proficiency by drivers, which in turn can likely be attributed to methodological errors in the experiment itself. These errors are cause for concern and will be also discussed in further detail in later sections.

Rest area task. In a task designed to assess the strategic level of driving performance, participants were given a description of a rest area and asked to pull into this rest area when they came to it. The task was based on the assumption that overloaded participants would fail to see the rest area and thus drive past it without entering. However, debriefings with participants revealed that many participants who failed to enter the rest area did in fact see it but were unsure if what they saw was the rest area in question. Thus, despite only half of all drivers entering the rest area, it was not possible to reliably attribute this to overload and analysis of this data could potentially be misleading.

Conversation Measures

Audio-visual recordings were made of all the experimental sessions and the 20 seconds of conversation surrounding each hazard (ten seconds before and ten seconds after) was transcribed for each participant dyad in the Passenger, Cell phone and Videophone conditions. All participants in each of the above groups were instructed to begin their conversation two minutes before they started driving and the 20 seconds of conversation immediately prior to when driving began was also transcribed and used as baseline data. Each conversation segment was then analysed based on three different discourse measures - mean utterance length, number of pauses, and percent of situation awareness utterances. A separate calculation of each measure was performed for drivers and conversers for each hazard, and this data was then combined to create a mean driving value for each measure. Again, small sample size dictated that statistical analyses be conducted separately for each conversation measure.

Mean utterance length was measured in number of words, and was the total number of words divided by the total number of utterances. 'Filler' words such as "um" and "ah" were included in the word count, and exclamations such as "Oh!" were also considered to be utterances. Laughter was excluded from utterance counts. Pauses were restricted to gaps in speech longer than two seconds that occurred mid-utterance, and were included regardless of whether or not the utterance was ever revisited. Pauses of longer than two seconds that occurred between utterances were considered to be products of normal turn-taking behaviour and were not included in this particular analysis. Instead, the total number of words spoken by both parties in the 20 second time frame was counted and used as a comparison, as differences in the number of pauses between utterances could potentially be reflected in this measure. Situation awareness utterances included any utterance in which the topic of conversation was the immediate driving task or situation, and this measure was calculated as a percentage of the total number of utterances made during the 20 second window.

Utterance length. Figures 16 and 17 show the mean utterance length for drivers and conversers for both baseline and whilst driving. Drivers showed a degree of variation between groups during baseline but this variation decreased when driving began, with all conversation conditions displaying similar means. Conversers showed variation between conditions during both baseline and whilst driving; however the direction of differences shifted as driving began, with the mean utterance length of conversers in the remote conversation conditions increasing as a result of driving while the utterance length of conversers in the Passenger condition decreased during driving.

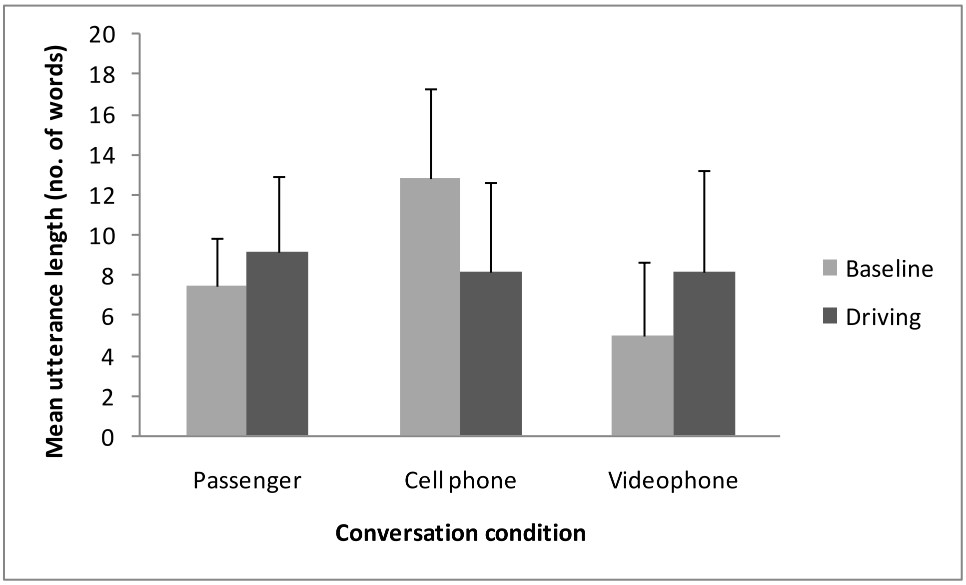


Figure 16. Mean utterance length for drivers by conversation condition, as exhibited during baseline and whilst driving. Error bars show 95% confidence intervals.

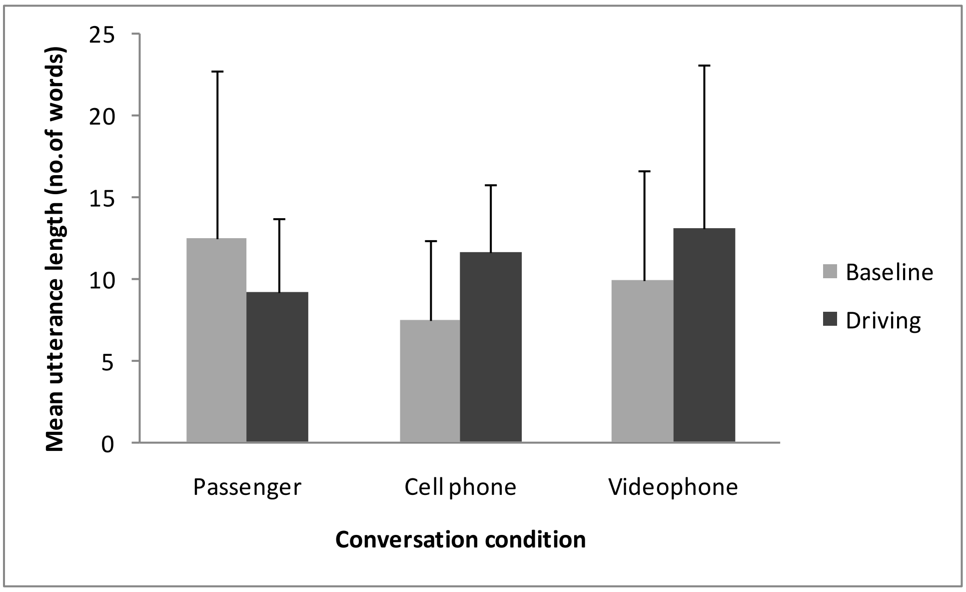


Figure 17. Mean utterance length for conversers by conversation condition, as exhibited during baseline and whilst driving. Error bars show 95% confidence intervals.

The drivers' mean utterance lengths were compared using a 3 x 2 mixed design multivariate analysis of variance, with three levels of the between-subject variable, conversation condition, and two levels of the within-subject variable, driving state. This analysis failed to reveal any significant differences between conversation conditions ($F_{(2,15)} = 2.425$, $p > 0.05$, $\eta_p^2 = 0.244$), but did reveal a significant interaction between conversation condition and driving state ($F_{(2,15)} = 4.647$, $p < 0.05$, $\eta_p^2 = 0.383$). Subsequent one way analyses of variance revealed the source of this interaction to be significant differences between the mean utterance length of drivers during baseline ($F_{(2,15)} = 8.105$, $p < 0.01$). It was also revealed that these differences disappeared when participants began driving ($F_{(2,15)} = 0.106$, $p > 0.05$), hence the presence of a significant interaction but no significant between-condition differences in the multivariate test. Post-hoc tests made using the Bonferroni adjustment revealed that drivers in the Cell phone condition made significantly longer utterances than drivers in both the Passenger and Videophone conditions during baseline (ps < 0.05 and < 0.01 respectively). A separate 3 x 2 mixed design multivariate analysis of variance was used to compare the mean utterance lengths of conversers and this failed to reveal any significant differences ($F_{(2,15)} = 0.216$, $p > 0.05$, $\eta_p^2 = 0.028$) or interactions ($F_{(2,15)} = 1.21$, $p > 0.05$, $\eta_p^2 = 0.139$).

Pauses. Figure 18 shows the mean number of pauses made by drivers and conversers in each conversation condition whilst driving. As only six pauses were made by both drivers and conversers in all conversation conditions across all five hazards, statistical analysis was unfeasible.

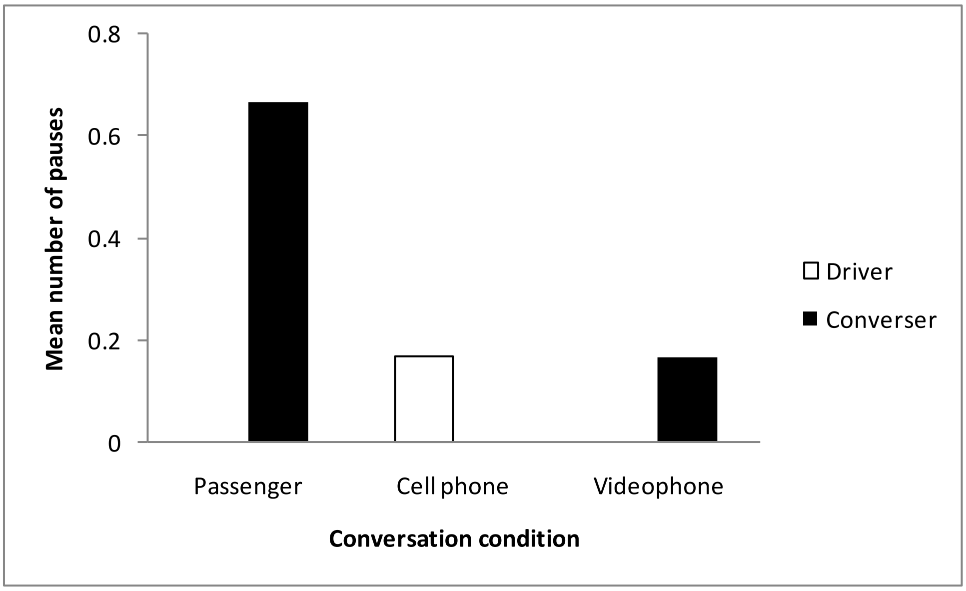


Figure 18. Mean number of pauses for drivers and conversers by conversation condition.

Number of words. Figure 19 shows the mean combined number of words spoken by drivers and conversers during baseline and whilst driving for each of the conversation conditions. As can be seen in the graph, there was little variation both within and between conditions, with statistical analyses revealing no significant within or between condition differences ($ps > 0.05$).

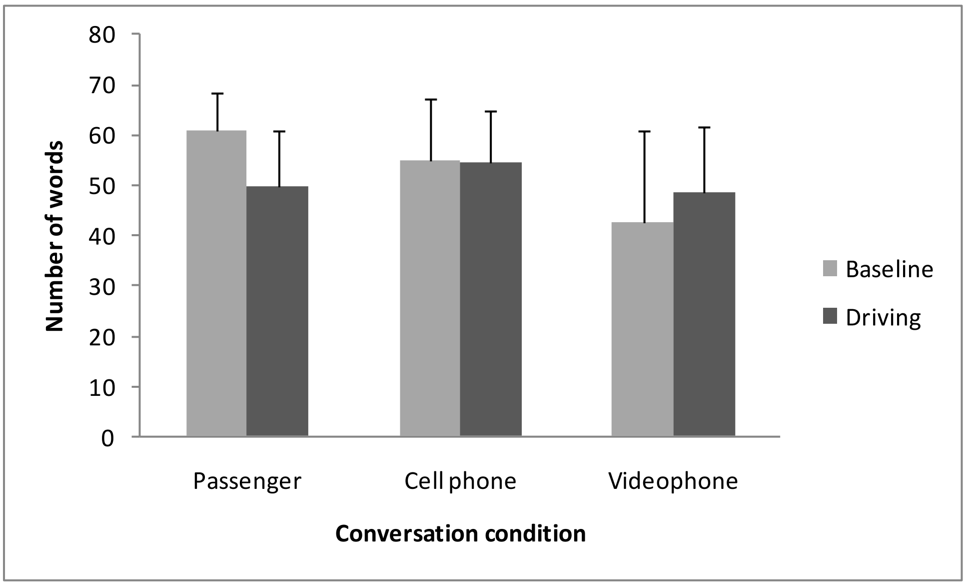


Figure 19. Mean total number of words spoken by participant dyads during 20-second segments for both baseline and whilst driving. Error bars show 95% confidence intervals.

Situation awareness utterances. Figure 20 shows the mean percent of situation awareness utterances made by drivers and conversers in each of the conversation conditions and indicates that both drivers and conversers in the Passenger condition made a higher percentage of situation awareness utterances than their counterparts in the Cell phone and Passenger conditions. Baseline data is not shown because given that a situation awareness utterance was any utterance relating to the immediate driving task, these utterances could not exist in baseline measures.

The figure also shows that in some cases (e.g., for conversers in the Videophone condition), the upper bound 95% confidence interval was more than double the group mean. Such large standard errors (in relation to the sample mean) strongly suggest the sample may not be a reliable representation of the population, another problem that will be addressed in the following chapter.

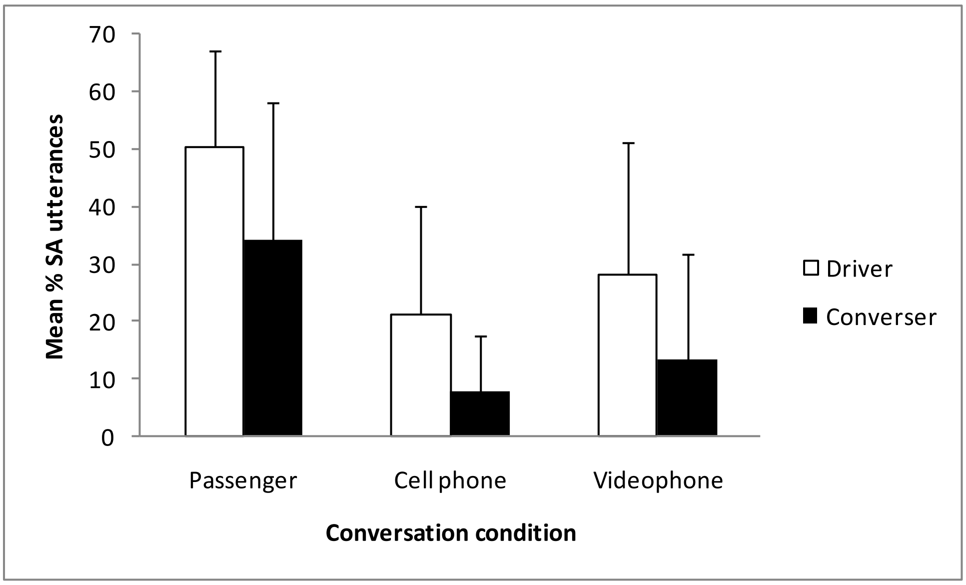


Figure 20. Mean percent of situation awareness utterances for drivers and conversers by conversation condition. Error bars show 95% confidence intervals.

The mean percentage of situation awareness utterances made by drivers and conversers in each conversation condition were compared using two separate one way analyses of variance, and significant differences were found for both drivers ($F_{(2,15)} = 5.206$, $p < 0.05$) and conversers ($F_{(2,15)} = 6.086$, $p < 0.05$) . Post hoc pair-wise comparisons made using the Bonferroni adjustment confirm that both drivers and conversers in the Cell phone condition made significantly fewer situation awareness utterances than their counterparts in the Passenger condition (ps < 0.05).

Difficulty and interference ratings

Drivers were asked to rate the difficulty of driving the simulated road on a 7-point scale, with 1 being "easy" and 7 being "impossible". Both drivers and conversers were asked individually to rate the difficulty of having the conversation using the same scale. Drivers and conversers were also asked to rate how much they felt the conversation interfered with the driving task on another 7 point scale, with 1 being "no interference" and 7 being "so much interference that driving was impossible".

The results of these ratings are shown in Figures 21-25. Conversation difficulty as rated by conversers was the only rating to reveal a significant difference between conditions ($F_{(2,15)} = 4.032$, $p < 0.05$), with post hoc comparisons (again made using the Bonferroni adjustment for experiment-wise error rate) revealing conversers in the Cell phone condition rated their conversations as significantly easier than those in the Videophone condition ($p < 0.05$). There was no significant difference between conditions for conversers' ratings of conversation interference ($F_{(2,15)} = 1.06$, $p > 0.05$). No significant differences were found for any of the driver measures (*ps* > 0.05). Interestingly, comments from participants suggested conversers in the Videophone condition seemed less confident in their ability to assess the level of interference experienced by drivers than did the conversers in the Cell phone condition, despite those in the Videophone condition having access to visual cues.

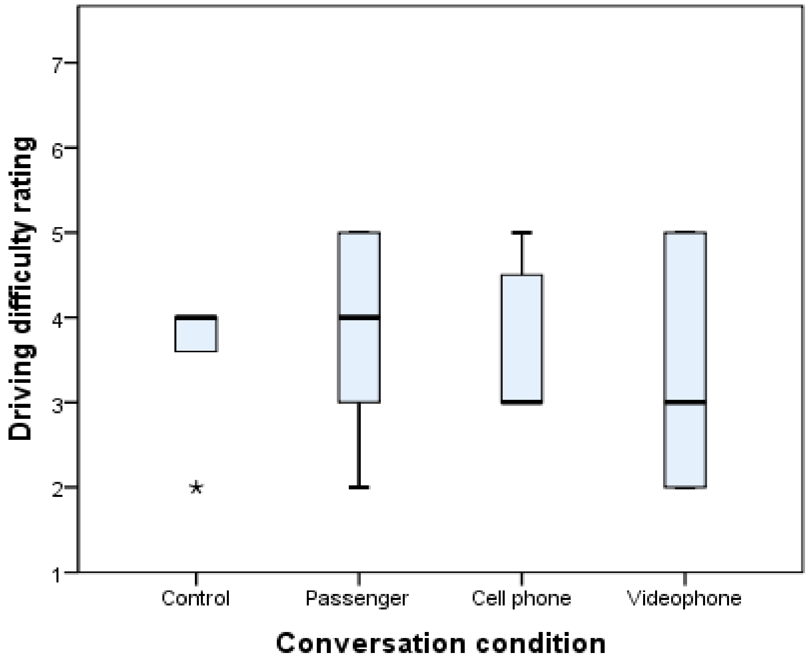


Figure 21. Drivers' ratings of driving difficulty, where 1 is "easy" and 7 is "impossible." Bold lines indicate the median, shaded areas indicate the inter-quartile range, and whiskers indicate the range. Asterisks indicate outliers, which include any value more than 1.5 times the inter-quartile range greater or less than the next closest score.

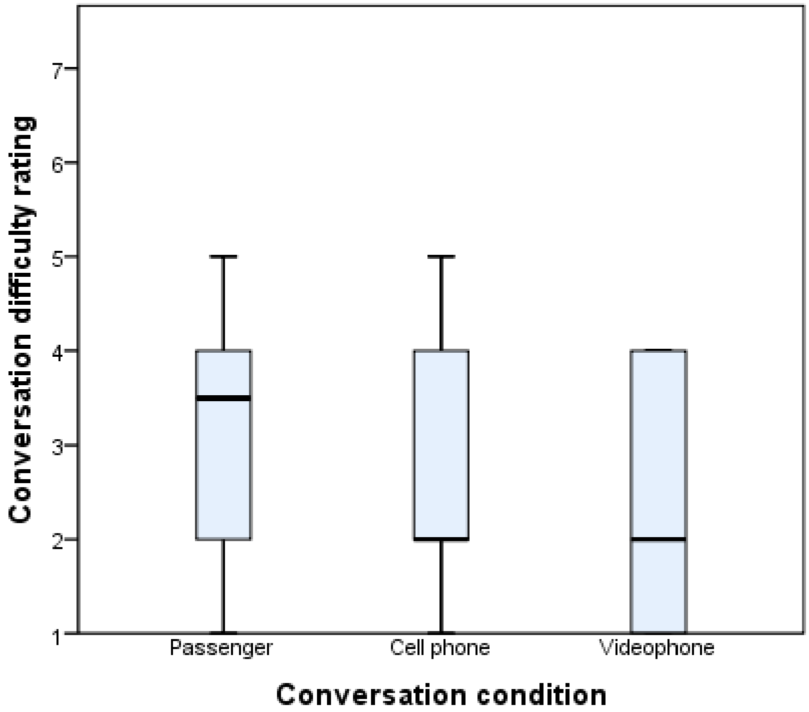


Figure 22. Drivers' ratings of conversation difficulty, where 1 is "easy" and 7 is "impossible." Bold lines indicate the median, shaded areas indicate the inter-quartile range, and whiskers indicate the range.

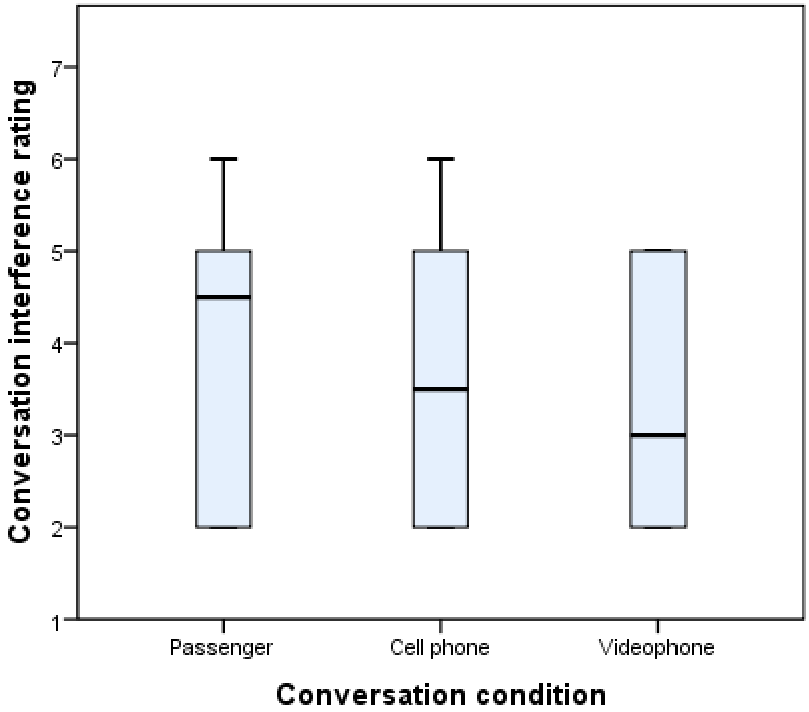


Figure 23. Drivers' ratings of conversation interference, where 1 is "no interference" and 7 is "so much interference that driving was impossible". Bold lines indicate the median, shaded areas indicate the inter-quartile range, and whiskers indicate the range.

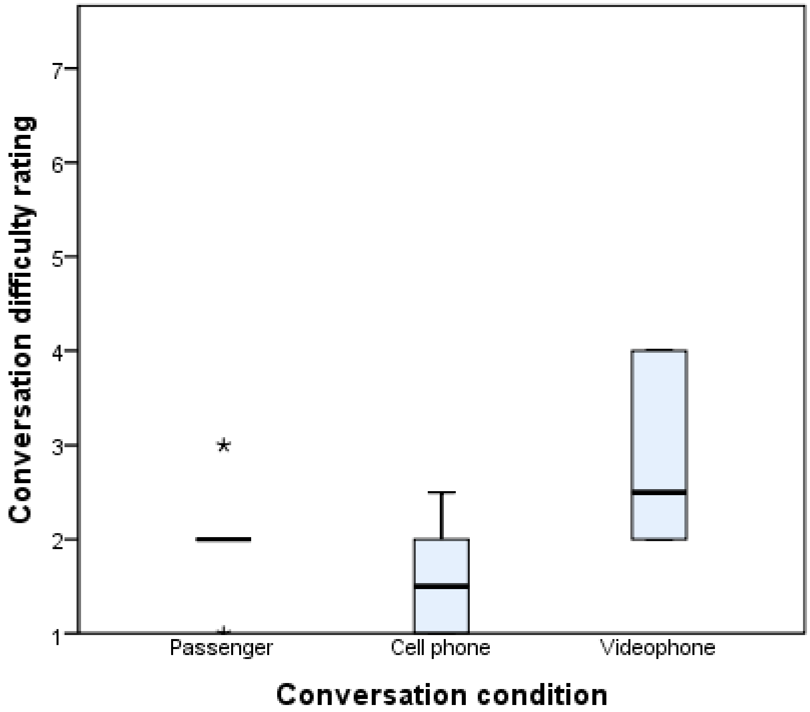


Figure 24. Conversers' ratings of conversation difficulty, where 1 is "easy" and 7 is "impossible." Bold lines indicate the median, shaded areas indicate the inter-quartile range, and whiskers indicate the range. Asterisks indicate outliers, which include any value more than 1.5 times the inter-quartile range greater or less than the next closest score. Note the maximum value of 2.5 in the Cell phone condition was due to a participant giving a rating of "2-3."

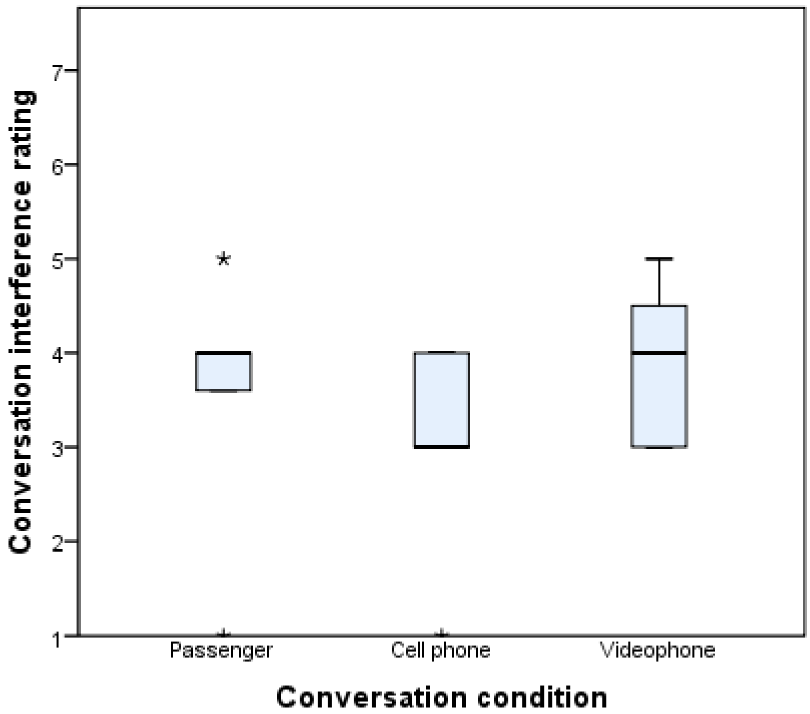


Figure 25. Conversers' ratings of interference experienced by drivers as a result of conversation, where 1 is "no interference" and 7 is "so much interference that driving was impossible." Bold lines indicate the median, shaded areas indicate the inter-quartile range, and whiskers indicate the range. Asterisks indicate outliers, which include any value more than 1.5 times the inter-quartile range greater or less than the next closest score.

Discussion – Experiment 1

One of the more noticeable aspects of the findings of this experiment was a significant difference between the number of situation awareness utterances made by those in the Passenger and Cell phone conditions. In accordance with the literature, conversers in the Cell phone condition made significantly fewer situation awareness utterances than those in the Passenger condition. One of the goals of this experiment was to investigate whether having access to visual cues would increase the likelihood of situation awareness utterances being made by either drivers or conversers. While there were no significant differences between the Videophone condition and the other two conversation conditions in this measure for either drivers or conversers, the Videophone condition showed particularly large within-group variability, to the point that the upper confidence interval was more than twice the group mean in the case of conversers. While this is usually a sign that the sample data is not a reliable representation of the population parameters and this must be considered as a potential limitation, it is also possible that for some reason, use of a videophone created noticeable differences in speech patterns in some individuals and not others, and this instead caused the large within-group variability. It could be that the sample in fact included members of two distinct populations – one for whom videophones had an effect and one for whom they did not. Investigating whether or not these two populations exist and if so, learning in what ways the members differ are potential areas for future research.

As highlighted in the previous section, several problems arose during the running of Experiment 1. One of the most notable problems to arise was an unexpectedly high crash rate. While conversation has been associated with an increase in crash rate (as discussed in the Introduction), there appeared to be no specific relationship between conversation condition and crash rate in this experiment, with participants in the Control condition crashing just as often as those in the other conversation conditions. Also of concern was the severity of the crashes. As opposed to minor incidents such as sideswiping another object, the majority of crashes involved participants' cars spinning more than 180° and often required the simulator to be reset. Several participants had as many as three of these serious crashes over a distance of less than 15 kilometres. The simulated road used in Experiment 1 was closely matched to an actual section of New Zealand road, located between Bethlehem and Katikati in the Bay of Plenty, and it seemed improbable that participants would experience as many crashes of a similar nature if they were to drive the actual road in their own car, whether or not they were engaging in conversation. Rather, it seemed that participants were not displaying an adequate level of proficiency in handling the simulator and as such, the driving behaviour exhibited in the simulator by participants was not an accurate reflection of normal driving behaviour.

This observation prompted preliminary analysis of the data, and this analysis revealed high levels of within-group variability in the majority of both driving and conversation measures. In some instances a specific cause could be identified, such as in the case of participants’ speeds at Hazard 3, in which poor road design allowed for large within-group differences as discussed in the previous section. However, for the majority of measures, no obvious cause of variability was present. Also perplexing was the high level of variability within individual subjects. For example, one participant drove 50km/h over the speed limit through one hazard then 10km/h below the speed limit through the next. It is possible that this could be attributed to the lack of proficiency displayed by drivers as mentioned above. It is also possible that the impact of this lack of proficiency may have been heightened by drivers' awareness of their own deficiencies in this area. For example, if a driver was aware that depressing the brake pedal too hard or too quickly may cause the vehicle to skid but they did not feel confident in their ability to depress the brake pedal in the correct way, they may have been more reluctant to brake, therefore resulting in higher speeds than normal in some cases. Conversely, this same awareness could have resulted in lower speeds if drivers tried to reduce the number of situations in which they would need to brake by driving at speeds lower than normal. This would also have implications for reaction time and time-to-collision, and could have contributed to an increased number of participants failing to react as they approached a hazard, despite needing to do so in order to negotiate the hazard safely.

Another factor of concern was the number of drivers who failed to respond to changes in speed limit. As mentioned in the Results section, the majority of crashes happened at Hazards 2 and 3. These hazards were both designed in such a way that if participants slowed to the sign-posted speed limit, they would be able to negotiate the hazard successfully without necessarily having to make a braking response. Yet not only were many of the participants crashing at these sites, but these crashes were particularly severe, as outlined above. This is reflective of the high speeds displayed by participants. While it is possible that this is yet another product of participants' general lack of proficiency in handling the simulator, it is also possible that participants exhibited such high speeds at hazards because they were actually unaware of the preceding drop in speed limit. To investigate this possibility, reactions to the speed limit signs preceding Hazards 2 and 3 were examined, and it was found that many participants failed to exhibit any deceleration response as they approached and passed the speed limit signs. (In both cases, the new speed limit was marked by speed limit signs on both sides of the road, identical to those used on real New Zealand roads). Again, while it is possible that this could be due to the participants' hypothesised fear of braking as outlined above, another likely possibility is that participants failed to respond because they did not perceive or process the signs as they passed them. If so, it could be assumed that this deficit in perception was the result of cognitive overload, but given that drivers in the Control condition were just as likely to miss speed limit signs as drivers who were conversing, it is also fair to assume that this overload was due to some factor other than conversation. Instead, it may be that the driving task (i.e. this particular simulated road and its hazards) was too difficult to produce the expected driving behaviour in participants.

Regardless of their respective causes, which cannot be conclusively determined without further empirical testing, the presence of the above combination of factors ultimately led to a decision to terminate Experiment 1 after running just 24 participant dyads, instead of the intended 64. Full analysis was then performed on all data collected prior to that decision. However, as can be expected when working with a data set of such limited size, this created several additional limitations of this study, which in turn had implications for interpreting the data correctly.

One such limitation involved the gender of participants. Participant dyads were randomly assigned to a conversation condition at the time of giving initial consent for involvement in the experiment. By chance, most of the male dyads were assigned to the Videophone condition. While this imbalance would likely have corrected itself had the full 64 participants completed the experiment, the early termination meant that at the time of analysis, the Control, Passenger and Cell phone conditions all included mostly female participants and only one or two male participants, if any; whereas the Videophone condition included just one female, with the remaining five drivers in this condition being male. Given the known behavioural differences between males and females, this has implications for both driving and conversation measures and must be acknowledged as a limitation of this study.

The small sample size also has implications for accurate interpretation of conversation data, considering the nature of normal conversation and the fact that this study examined drivers' and conversers' data separately. During the course of a conversation, it is normal for one party to "hold the floor" for a time, perhaps as they tell a story, meaning this party makes longer utterances while the other may make only one-syllable utterances, as a means of letting their partner know they are listening. This pattern may then reverse several times throughout the conversation, meaning that if the whole conversation were to be examined, each party would have a similar mean utterance length. However, if only a small number of short conversation segments were extracted from various stages during the conversation and examined as was the case in this experiment, it is possible that, by chance, the same speaker may have been leading the conversation each time, artificially inflating the utterance length of that party while deflating the other and potentially increasing the likelihood of making a Type I error. The statistical analysis for this study revealed a significant difference in utterance length between different conversation conditions during baseline for drivers only. Considering the baseline data set for each condition was based on only one 20-second segment from each of the six participant dyads, it is possible that, by chance, during those particular 20-second windows the conversers in the Cell phone condition were leading the conversation while the driver listened, while the reverse was true for participants in the Passenger and Videophone conditions. One possible way of determining whether or not this was in fact a Type I error could be to increase the size of the data set by using a longer conversation segment from each participant dyad to create the baseline data. While this was a possibility given that participants began their conversations two minutes before driving began and this full two minutes was captured in the audio-visual recordings, due to the number of other factors that may also be affecting the results of this study, it did not seem sensible to undertake further analysis but rather it seemed best to just acknowledge this as yet another limitation of this study.

It is also worthwhile to consider that the reverse of above phenomenon may also be true and could serve to explain why the significant difference in utterance lengths between conversation conditions disappeared when driving began. If chance dictated that the converser was leading the conversation as the driver passed through one hazard, but the driver was leading the conversation as they passed through the next, this may have obscured any changes in conversation form in response to the road, resulting in a Type II error. One obvious way to reduce the likelihood of this is to ensure an adequate sample size, but as this was obviously not possible in this case given the early termination of the experiment, this too becomes a limitation.

This pattern repeats itself when each of the other measures is examined. Thus, all of the results in this study should be interpreted with caution, with the exception of situation awareness utterances. The likelihood of someone making an utterance regarding changes in the driving situation by chance is almost nil; therefore this stands as perhaps the only result of value to emerge from this study.

Given this unexpected outcome, a decision had to be made regarding the future direction of this research project as a whole. Upon review, it appeared two options were available. The first option was to abandon the original research question to empirically pursue the reasons for the unexpected results obtained in Experiment 1. The second option was to continue pursuing the original research questions by creating a list of possible (but untested) reasons for the unexpected outcomes, designing countermeasures for each and then incorporating each of these countermeasures into a redesigned version of the experiment. While both options may have produced interesting findings, the decision was made to pursue the original research questions through a second experiment, and this experiment forms the basis of the second half of this thesis.

Introduction – Experiment 2

Although the literature indicates the potential existence of differences between the effects of various conversation modes on driving performance, the unexpected results obtained in the previous experiment suggested a number of methodological issues were hindering the observation of any such effects. In order to address these issues and enable continuation of the research, potential causes of each of the problems encountered during Experiment 1 were identified and rectification strategies for each were produced and incorporated into a modified experimental design, as outlined below.

One of the major problems present in Experiment 1 was a high level of within-group variability. It appeared the most appropriate methodological change to counter this would be a switch from a between-subjects to a within-subjects design. While a within-subjects design does not reduce the natural variability between participants that may likely have also contributed to the high within-group variability in Experiment 1, it limits the impact of any between-subject differences and allows for more accurate interpretation of the data. Under the new experimental protocol, drivers completed the experimental drive four times, once under each of the four conversation conditions. In order to limit the strain placed on participants, the experimental drive was shortened from 25km to 10km and included just three hazards. Each drive took approximately eight minutes and the entire experiment was able to be completed within an hour.

It is important to be aware that the use of a within-subjects design introduced the possibility of new methodological issues arising, and these also needed to be considered and addressed to ensure a successful outcome in Experiment 2. One such issue was the possibility of learning effects. Given that drivers drove the same road four times within one session, with exactly the same hazards, it is reasonable to expect better (or at least more stable) driving performance in the later trials, especially amongst novice drivers. It was also possible that the reverse of this pattern could occur, if drivers became so familiar with the task that they inadvertently devoted an inadequate amount of attention to the driving task in later trials. To combat the effects of both of these potential outcomes, participants completed the trials in counterbalanced order.

Another possibility that needed to be considered was that of order effects, not as a result of practice but rather as a result of the effect of the previous conversation condition. For example, a driver completing the Passenger condition trial immediately after the Control condition trial may assign much higher difficulty and interference ratings to the Passenger condition than a driver who completed the Passenger trial after the Cell phone trial. Not only may difficulty and interference ratings be affected, but it is also possible that the previous conversation method may have a lingering impact on the cognitive state of the driver and therefore also affect driving and conversation performance. Thus, in order to control for these possibilities, rather than using a simple Latin square as the basis for the counterbalanced order, in which the order of trials is adjusted sequentially with the same conditions preceding and following each other in most cases, a special order for participants was derived. The basic rules of Latin squares were retained, with each conversation condition appearing an equal number of times in each row and column, however it was arranged in such a way that each conversation condition preceded each of the other three conversation conditions an equal number of times across the full experiment. The table for the order for participants is included in the appendices.

Although a within-subjects design introduces the possibility of new methodological issues arising as outlined above, it also introduces a number of additional benefits. One such benefit is that it allows for a larger data set with fewer participants, which in turn reduces the likelihood of the statistical issues that arose from the use of small data set in Experiment 1 occurring in Experiment 2, even though the number of driving participants was unchanged. Another benefit of a within-subjects design is that participants act as their own controls, eliminating the need for baseline conversation data (which was subject to interpretation issues in Experiment 1). The switch to a within-subjects design also eliminated the risk of any gender imbalances within the conversation conditions as found in Experiment 1.

Another of the major problems encountered during the running of Experiment 1 that needed to be addressed was the drivers' general lack of proficiency in handling the simulator. There are many factors that could have contributed to this lack of proficiency. If one assumes that the drivers involved in Experiment 1 came from a normal population and possessed an adequate level of proficiency for real-world driving, the next logical step could be to assume that the fault lay in the simulator itself. However, given that other researchers have successfully conducted many other studies using the University of Waikato simulator, this did not appear to be a logical assumption. Instead, it may have been that the poor handling shown by participants was in some way due to interference arising from the subtle differences between the simulator and the participant's own car, a phenomenon not unique to simulators but which is also present when someone drives a new car for the first time. To combat this, it was decided to include a group of practiced participants, who had each accrued at least ten hours of driving experience in the University of Waikato simulator during a separate study. The inclusion of this group is perhaps the most prominent change between Experiment 1 and Experiment 2 and has much potential for producing interesting findings.

It is also worthwhile to consider the role of the experimenter when attempting to identify potential causes of the lack of proficiency displayed by drivers in Experiment 1. It is possible that there was some form of negative experimenter influence present during the trials. The experimenter's inexperience in teaching people to drive the simulator may have somehow been projected onto participants, instilling a lack of confidence in drivers which in turn may have negatively affected their driving performance. Alternatively, it may have been that the experimenter did not exhibit enough of an authoritative manner, and as a result drivers did not feel the need to treat the simulation as real. In order to control for this and any other potential experimenter influence, a second experimenter was introduced in Experiment 2. The experimenters ran their assigned sessions independently, allowing for easy comparison between the data collected by each experimenter. Advice on good teaching techniques for the simulator were sought from other University of Waikato researchers and these were incorporated into a revised set of operating instructions which each experimenter used when teaching new participants to operate the simulator. These are also included in the appendices.

The lack of proficiency displayed by drivers also called for a review of the practice given to participants. Even if participants were to experience some interference as a result of the differences between the simulator and their own car, this should be able to be reduced to a manageable level with sufficient practice. Experiment 1 included a 10-15 minute practice run at the start of each session, yet this was not sufficient to ensure adequate driving performance. However, rather than making the practice longer, it was decided that the best course of action was to endeavour to make the practice more efficient. As a result, a new practice route was created for Experiment 2. The practice route was exactly the same as the experimental road, but with the hazards removed. Traffic was kept at the same rate as it appeared in the experimental sessions, and a set of road works were included in the practice to show participants an example of the type of hazard they could expect to see during the experimental drive. Participants were made aware that the practice road was the same road they would be driving in the experiment (minus the hazards), and as participants drove the practice road, the experimenter talked them through the various road features, taking care to point out changes in speed limit and the rest area located at the end of the drive. It was hoped giving drivers this knowledge would address the problem with identification of the rest area that occurred in Experiment 1, and also allow for a more reliable attribution to cognitive overload should a participant fail to respond to a speed limit sign. Also, given that the practice and experimental roads were almost identical, a successful practice drive assured the experimenter that the participant possessed enough skill to complete the experimental trials successfully.

There was also evidence to suggest that the level of difficulty of the driving task used in Experiment 1 was too great, resulting in the drivers experiencing cognitive overload, which in turn negatively affected driving behaviour and made interpretation of conversation data difficult. As a result, the initial speed limit was reduced from 100km/h to 60km/h for the first two kilometres to allow time at the start of each drive for participants to reacquaint themselves with the subtle differences between the simulator and a real car if necessary. This meant Hazard 1, in which drivers had to negotiate a tight gap between a police car and turning truck, was now located within this slower speed zone, allowing participants more time to process and respond to the hazard. To increase simplicity, the speed limit was kept at 100km/h for the remainder of the drive, before dropping to 60km/h just prior to the rest area to alert the driver the end of the drive was approaching and to allow safe and easy entry into the rest area. This new speed limit configuration meant Hazard 2 was now located within a 100km/h zone, and it was hoped this change would reduce variability in participants' approach speeds at this hazard, as much of the within-group variability at this hazard in Experiment 1 could be attributed to some participants failing to notice the 70km/h sign preceding the hazard. Also, due to the high crash rate observed at Hazard 2 in Experiment 1, the timing of the car pulling out was adjusted beyond what was necessary to accommodate the above change in speed limit so that the parked car entered the road much earlier, giving participants more time to react and reducing the chances of a collision.

The other troublesome hazard, the one lane bridge, was eliminated entirely and replaced with a set of road works, a hazard which all participants managed to successfully negotiate in Experiment 1. The speed limit temporarily dropped to 30km/h briefly at the road working site before returning to 100km/h.

One problem that has so far not been mentioned and is slightly difficult to address specifically is that of the high levels of variability displayed within individual subjects in Experiment 1. However, as outlined in the previous section, it is possible that this variability was connected to the participants' lack of simulator-handling proficiency and possible cognitive overload; therefore given that both of these issues have been addressed previously, the inclusion of the above countermeasures should lead to a reduction in this particular type of variability.

While it was possible that these countermeasures may not have been completely sufficient in addressing the issues encountered during the running of Experiment 1, they formed a good basis for a well-controlled experiment and it was hoped that they would greatly increase the likelihood of any differences between the effects of passenger, cell phone, and video phone conversations on driving performance being detected, should they exist.

Method – Experiment 2

Participants

Experiment 2 included both novice and practiced drivers and participants for these two groups were recruited in different ways. The Novice group included both University of Waikato students and members of the general public, and these participants were recruited in the same way as participants recruited for Experiment 1. University of Waikato students were recruited via notices on both physical and electronic University of Waikato notice boards, and members of the general public were recruited via notices placed on community notice boards and through social networking sites. As per Experiment 1, participants were recruited in pairs with a friend or acquaintance to ensure fluent conversation.

The Practiced group consisted of drivers with a minimum of ten hours driving experience in the University of Waikato simulator. As such, potential participants for this group were limited to people who had been involved in long-term driving experiments run by other University of Waikato researchers. Upon consent for release of contact details being obtained by the supervisors of these long-term driving experiments, potential participants were contacted directly and those who agreed to participate then recruited a friend or acquaintance to join them as their conversation partner.

The Novice group included six male and six female dyads, with a mean age of 27.46 years and with a mean of 12.45 years of driving experience. Novice drivers drove an average of 265.83km per week while novice conversers drove an average of 374.25km per week. The Practiced group included nine female dyads and 3 male dyads, with a mean age of 34.79 years and a mean of 16.91 years of driving experience. Practiced drivers drove an average of 235.83km per week while practiced conversers drove an average of 154.17km per week. Seventy-seven percent of participants were NZ European, 10% were of Maori descent, with the remaining 13% including participants of Chinese, Fijian Indian, Dutch, English and European descent. All participants were fluent in English and all participants held a full licence except one converser in the Practiced group, who still held a Restricted licence.

The simulation scenario

The apparatus used in Experiment 2 was identical to that used in Experiment 1. The road used in Experiment 2 was also identical to that used in Experiment 1; however the road was shortened to include only the first ten kilometres. Only three of the five hazards used in Experiment 1 were retained in Experiment 2. These included the busy intersection with turning police car (Hazard 1); the parked car pulling out from beside a roadside pub (Hazard 2); and the road works (Hazard 3). Hazard 1 was not moved from its original location of 2.48km beyond the starting point. The location of Hazard 2 also remained unchanged, with participants encountering this 4.5km into their drive. The road works, which were in fact the fourth hazard encountered by participants in Experiment 1, were brought forward to a new location of 7.85km from the starting point, and became Hazard 3. The one-lane bridge that had served as Hazard 3 in Experiment 1 was removed and the road was widened to two lanes in this spot. A new rest area was added at the 10km mark, and participants were alerted to its presence by a sign 300m prior, and another at its entrance. This rest area marked the end of the drive.

Speed limits were also adjusted for Experiment 2. The initial speed limit was reduced from 100km/h to 60km/h for the first two kilometres, which resulted in Hazard 1 being located in a 60km/h zone in Experiment 2, as opposed to a 100km/h zone in Experiment 1. To account for this, the speed of the two heading vehicles was reduced from 96km/h to 52km/h. This configuration was exactly the same as that used by Charlton (2009). A 100km/h speed limit was introduced approximately 200m after Hazard 1, and was marked by a sign on the left side of the road. This meant Hazard 2 was now located in a 100km/h section, and the trigger for Hazard 2 was brought forward 80 metres, affecting the timing of the car pulling into traffic, as described in the previous section. With the exception of the 190m section of road surrounding the road works, where the speed limit temporarily dropped to 30km/h, the speed limit remained at 100km/h for the remainder of the drive, until approximately 650m before the rest area. Here the speed limit was once again reduced to 60km/h and this change was marked by signs on both sides of the road. No other changes to the location or speed of traffic were made, with the same mix of cars, light trucks and heavy vehicles being retained to maintain a rate of 8000-10000 passenger car units per day. Figure 26 shows the layout of the simulation scenario used in Experiment 2.

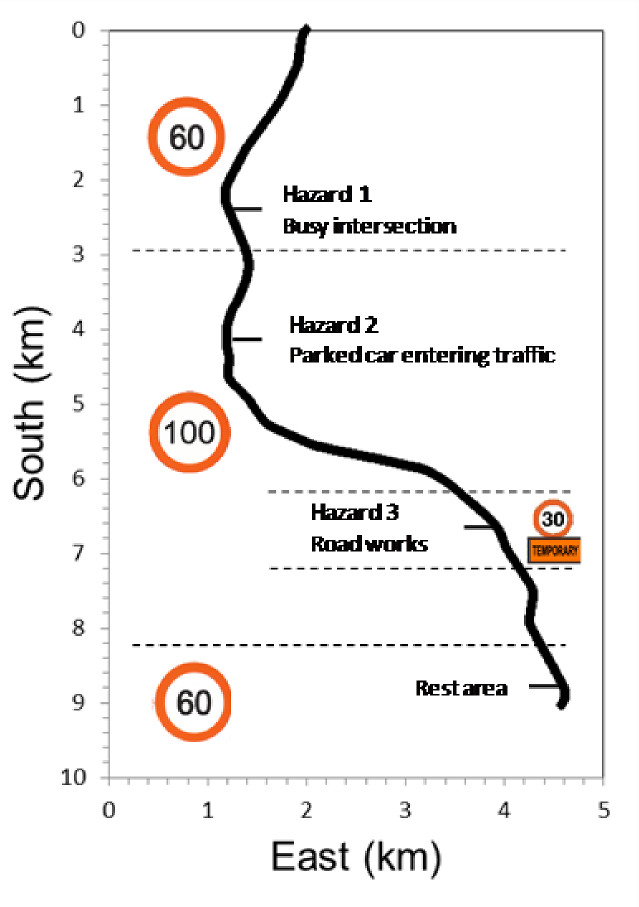


Figure 26. Map of experimental road used in Experiment 2, showing hazard placement, rest area and speed zones.

Procedure

Experiment 2 was a within-subjects design and each dyad completed each of the four conditions in counter-balanced order. The set-up for each condition was identical to that in Experiment 1. The sessions for Experiment 2 began in the same way as those from Experiment 1, with both participants filling out an informed consent form and a brief questionnaire about their driving and talking habits following a brief introduction by the experimenter. In the case of novice participants, each participant took a short drive before self-selecting who would be the driver for the experiment. This step was omitted when running practiced participants, as the person with ten hours' driving experience in the simulator was automatically given the role as driver. Once selected, drivers were taken through the new practice drive before beginning the experiment. Each drive lasted approximately eight minutes, and was followed by a short break. At the conclusion of each of the four experimental drives, participants were asked for difficulty and interference ratings.

Data collection and analysis

As per Experiment 1, each of the hazards served as a data collection point and speed and reaction times were collected at each one. The 20 seconds of conversation surrounding each hazard was transcribed and the total number of words, mean utterance length and percent of situation awareness were determined. Baseline conversation data was not gathered as subjects acted as their own controls. The data collected was analysed using mixed design multivariate analyses, as described in the next section.

Results – Experiment 2

Driving performance

Twenty-four drivers completed four trials each, meaning the data set used in this analysis was based on a possible 96 trials. Technical issues with the simulator resulted in the loss of the driving performance data for four of the 96 trials. Each of the four missing data sets belonged to a different participant, meaning no individual participant was missing data for more than one of the four trials they completed. In each case, the missing driving performance data were replaced with the condition mean for either the Novice or Practiced group, as appropriate.

Speed. As per Experiment 1, participants' speeds were observed at a series of successive location points surrounding each of the three hazards used in Experiment 2. Given the differing nature of the hazards, driving performance at each hazard was analysed independently, and the speed results for each of the hazards are described below.

Hazard 1. Figure 27 shows the mean speeds of drivers in the Novice group under each of the four conversation conditions as they approached and passed by Hazard 1, which consisted of a tight gap between a police car and a truck, both waiting to turn across the participant's path, but from different directions. Figure 28 shows the mean speeds displayed by drivers in the Practiced group under each of the conversation conditions as they passed this hazard. At this particular hazard, one driver in the Practiced group drove 40km/h above the posted 60km/h speed limit under every conversation condition except

Control, and as such, was an outlier at many of the measured location points surrounding this hazard (with an outlier being defined as any value that is more than 1.5 times the inter-quartile range greater or less than the next closest value). While it could be argued that the lack of excessive speed displayed by this particular driver in the Control condition suggests that their behaviour in the other conditions was a result of the distracting effects of conversation and should therefore be included in the analysis, the magnitude of this driver's increase in speed when driving and conversing at this hazard was great enough to potentially alter the outcome of the statistical analysis. As such, the participant's data for the Passenger, Cell phone and Videophone conditions were removed and replaced with the corresponding Practiced group means for each of these conversation conditions.

The figures show that at this particular hazard, both novice and practiced drivers exhibited the fastest speeds when conversing on a cell phone. No clear pattern emerged in the speeds displayed by drivers under any of the other three conversation conditions, either within- or between-subjects. Some conditions led to a decrease in speed as drivers approached the hazard while others resulted in participants speeding up as they neared the hazard; however these tendencies were not consistent across groups. For example, drivers in the Novice group slowed as they approached the hazard when driving under the Passenger condition, but increased their speed as the approached the hazard when driving under the Videophone condition. Practiced drivers, however, exhibited the opposite trend for these same two conversation conditions.

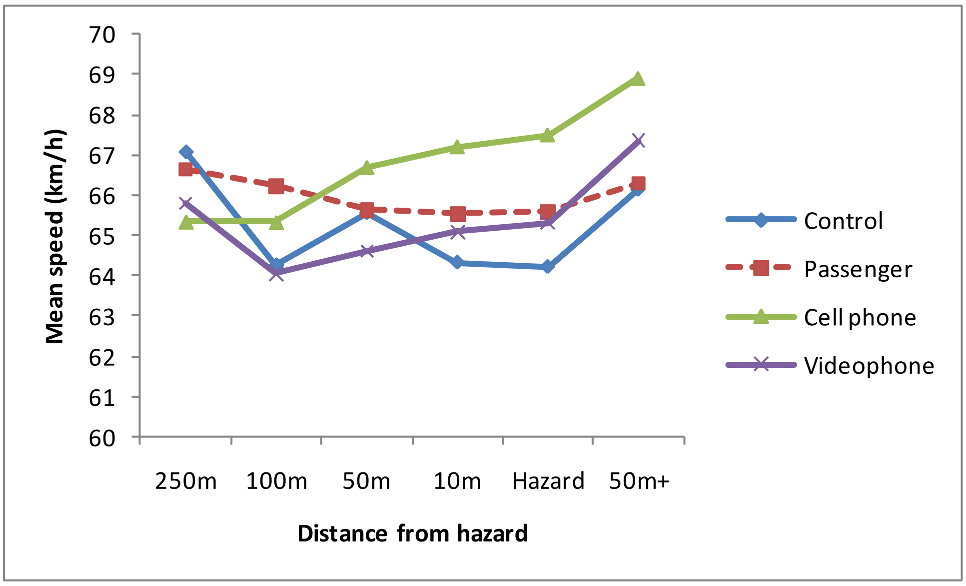


Figure 27. Mean speeds of novice drivers at Hazard 1, by conversation condition.

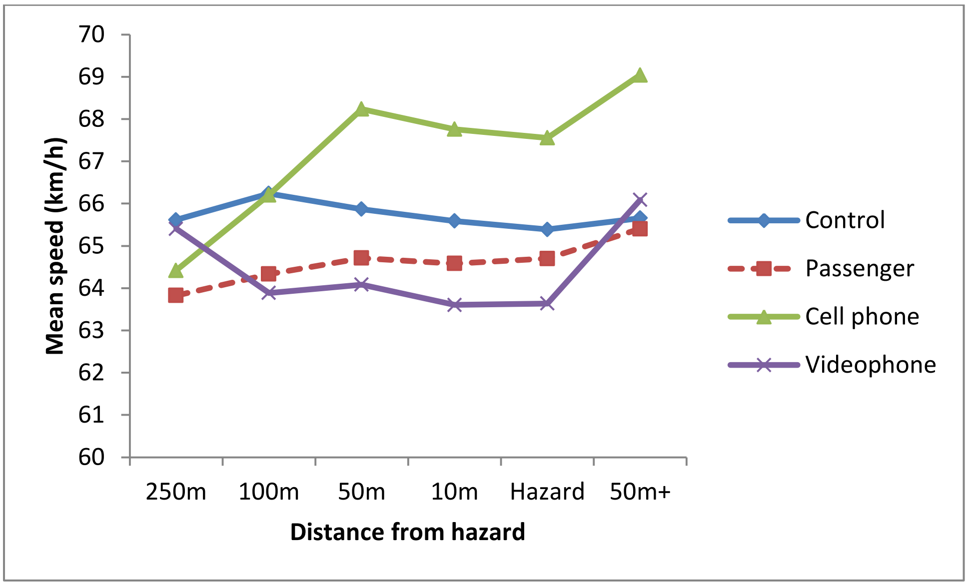


Figure 28. Mean speeds of practiced drivers at Hazard 1, by conversation condition.

A 2 x 4 mixed design multivariate analysis of variance was used to compare participants' speeds at Hazard 1, with two levels of the between-subject factor, practice (also referred to as "group" as participants were divided into two groups based on their level of simulator experience), and four levels of the within-subject factor, conversation condition (General Linear Model, PASW Statistics 18). As indicated by the graphs, this analysis failed to reveal any significant condition effect ($F_{(3,66)} = 1.818$, $p > 0.05$, $\eta_p^2 = 0.076$), practice effect ($F_{(1,22)} = 0.47$, $p > 0.05$, $\eta_p^2 = 0.002$), or any significant interaction between level of practice and conversation condition ($F_{(3,66)} = 0.359$, $p > 0.05$, $\eta_p^2 = 0.016$). Additional multivariate analyses of variance, identical in design, were used to compare the mean speeds of participants at each of the measured location points surrounding Hazard 1 and these also failed to reveal any significant effects of any kind (*p*s > 0.05).

Hazard 2. The mean speeds displayed by drivers at Hazard 2 are shown in Figures 29 and 30, with Figure 29 showing the means for drivers in the Novice group for each of the conversation conditions, and Figure 30 showing the same for the Practiced group. The previously-mentioned participant from the Practiced group also drove at excessive speeds in the Passenger and Cell phone conditions at Hazard 2, but produced relatively normal data in the Control and Videophone conditions. As such, the participant's Passenger and Cell phone data were replaced with the appropriate group means.

A number of other participants also produced outlier values at this particular hazard, but these were a result of the participants driving much slower than the other participants for the given conversation condition and group. Closer investigation revealed that six of the 24 participants failed to respond to a 100km/h sign introduced shortly after Hazard 1 during one of their four trials. Participants were deemed to have missed the speed limit change if, 250m beyond the 100km/h sign, they were still travelling within 10km/h of the speed they were doing as they passed Hazard 1 (located with the slower speed zone). The change in speed limit saw the speed limit increase by 40km/h, therefore a difference of less than 10km/h several hundred metres later is likely a result of natural fluctuations in speed as opposed to participants merely taking their time to speed up. Of the six participants that fell into this category, three were from the Novice group and three were from the Practiced group. This phenomenon occurred twice in the Passenger condition, three times in the Cell phone condition, once in the Videophone condition, and never in the Control condition. No participant failed to respond to the change in speed limit more than once, although they did not necessarily miss the speed limit sign during their first trial. Given the possibility of these outliers also impacting the results, the data for these participants were removed for the affected trials only, and replaced with the appropriate means for the conversation condition and group.

Figures 29 and 30 show an overall downwards trend in speed as drivers approached Hazard 2, irrespective of group or conversation condition. This is indicative of the nature of the hazard, in which a parked car entered the road slightly ahead of the participant, with most participants slowing to accommodate the vehicle. A condition effect emerges when comparing the Control condition with the two remote conversation conditions (Cell phone and Videophone). Both novice and practiced drivers drove faster in the Control condition than in the two remote conversation conditions, which resulted in similar speeds for both groups. A practice effect is also noticeable, with practiced drivers exhibiting much faster speeds than the novices for each of these three conditions.

The Passenger condition, however, produced a very different pattern of results. Not only does the practice effect disappear, with drivers in the Novice group actually driving marginally faster than those in the Practiced group under this condition, but a practice by condition interaction can also be observed. At Hazard 2, drivers in the Novice group drover faster in the Passenger condition than in any other condition, whereas the opposite was true for practiced drivers, who exhibited the slowest speeds when in the Passenger condition at this particular hazard.

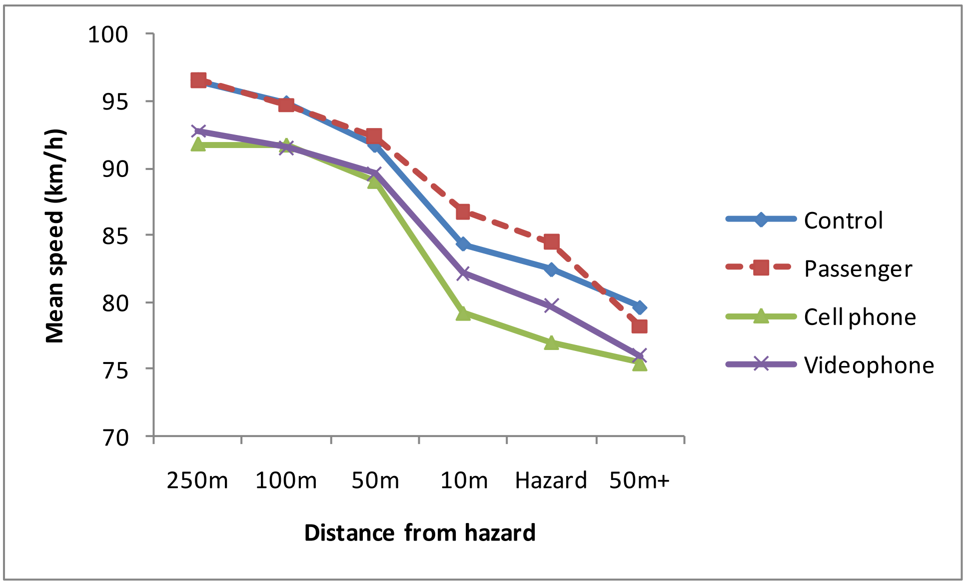


Figure 29. Mean speeds of novice drivers at Hazard 2, by conversation condition.

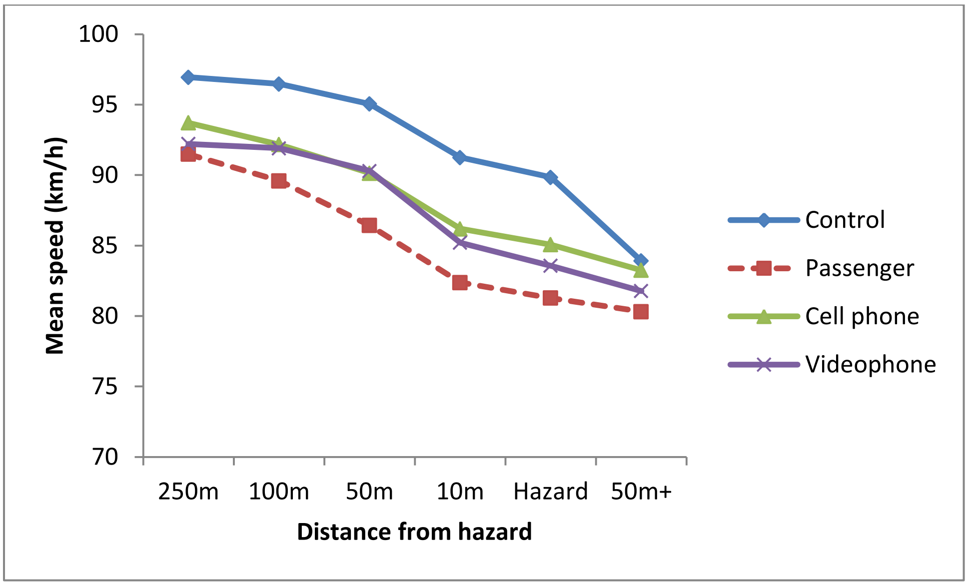


Figure 30. Mean speeds of practiced drivers at Hazard 2, by conversation condition.

Another series of 2 x 4 mixed design multivariate analyses of variance, identical to those used in Hazard 1, were used to compare participants' mean speeds at Hazard 2 itself and the five surrounding pre- and post-hazard location points. These analyses revealed a combination of significant within- and between-subjects effects and interactions. The condition effect described earlier, in which drivers drove faster in the Control condition than in the remote conversation conditions, was found to be significant 250m prior to the hazard ($F_{(3,66)}$ = 2.817, p < 0.05, η_p^2 = 0.114), and approached significance 10m prior to and at the hazard itself (10m prior: $F_{(3,66)}$ = 2.730, p = 0.051, η_p^2 = 0.110; at Hazard 2: $F_{(3,66)}$ = 2.664, p = 0.055, η_p^2 = 0.108). While none of the post hoc pair-wise comparisons made using the Bonferroni adjustment method reached significance, the pattern of results suggested a slightly stronger difference between the Control and Cell phone conditions than Control and Videophone.

The practice effect, in which drivers in the Practiced group drove faster than their Novice group counterparts in every conversation condition except Passenger, was found to be significant 50m beyond the hazard only ($F_{(1,22)}$ = 5.789, $p < 0.05$, η_p^2 = 0.208), but approached significance at the hazard itself ($F_{(1,22)}$ = 4.033, p = 0.057, η_p^2 = 0.135). A significant interaction, most likely a result of the differences in behaviour displayed by the two groups when driving with a passenger as explained above, was found 10m prior to the hazard and at the hazard itself (10m prior: $F_{(3,66)}$ = 3.978, $p < 0.05$, η_p^2 = 0.153; at Hazard 2: $F_{(3,66)}$ = 3.429, $p < 0.05$, η_p^2 = 0.155).

Hazard 3. Figures 31 and 32 show the mean speeds of participants at Hazard 3, a 190m stretch of road works involving loose gravel and a series of road cones. A 30km/h speed limit was introduced 30m before the hazard began. The same participant from the Practiced group once again produced some extreme data, this time driving 90km/h above the speed limit through the road works when in the Passenger condition only. As before, these outliers were replaced with the appropriate group mean. The participant's data for the remaining three conversation conditions were normal and were therefore included in the analysis.

As was the case in Hazard 2, the figures show an overall downwards trend in speed as participants approached the hazard, irrespective of group or conversation condition. In both groups, drivers produced the fastest speeds when in the Videophone condition and the slowest when in the Control condition. The Cell phone condition results also showed consistency between groups, falling between the Videophone and Control conditions for both novice and practiced drivers. However, just as was the case at Hazard 2, the level of practice seemed to have an effect on drivers' behaviour when driving with a passenger, but to a slightly lesser degree. Intriguingly, the effect appears to have reversed direction at this hazard, with novice drivers driving slower in the Passenger condition in comparison to other conversation conditions, while for those in the Practiced group, the Passenger condition produced some of the faster speeds.

It is important to be aware that given the large range of values shown in these particular figures, (most condition means decreased by approximately 50 km/h over the 300m measured due to the nature of the hazard) it is difficult to determine from observation of the figures alone whether any of these differences are indicative of actual effects or merely natural variation in the data. A clearer understanding can be obtained by examining the results of the statistical analysis.

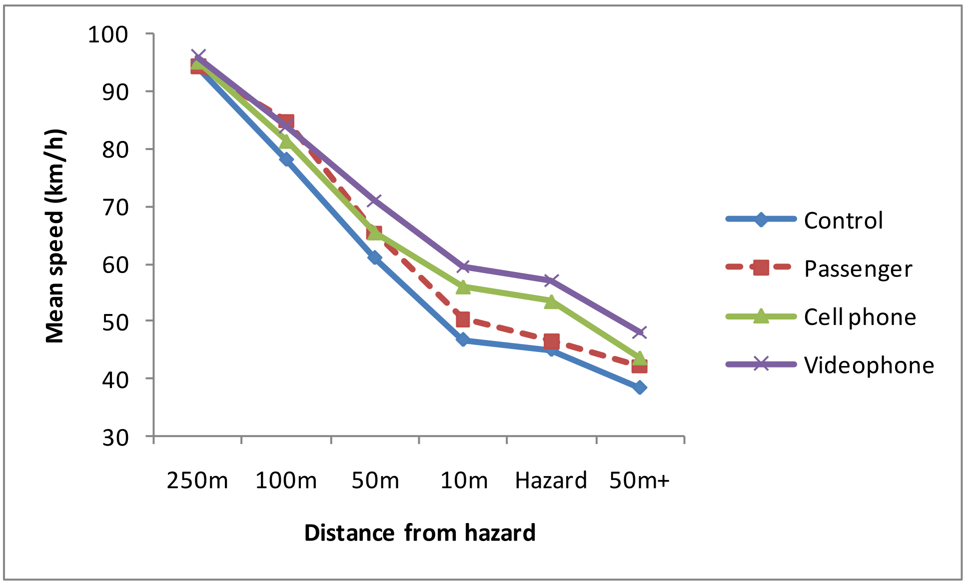


Figure 31. Mean speeds displayed by novice drivers at Hazard 3, by conversation condition.

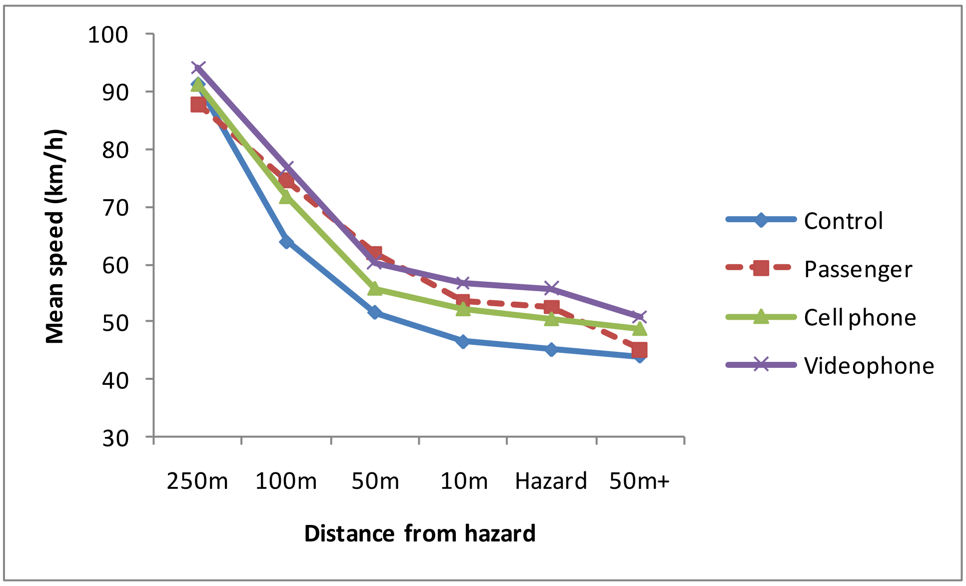


Figure 32. Mean speeds displayed by practiced drivers at Hazard 3, by conversation condition.

As with the previous two hazards, a series of six 2 x 4 mixed design multivariate analyses of variance were used to compare differences in speeds at this hazard. The results revealed there was in fact a significant condition effect at Hazard 3 ($F_{(3,66)} = 5.617$, $p < 0.01$, $\eta_p^2 = 0.203$), with post hoc pair-wise comparisons revealing participants drove significantly slower when in the Control condition than when in either of the two remote conversation conditions (*p*s < 0.05). It is interesting to note that this is the exact opposite effect to that found at Hazard 2. This condition effect was also significant at all of the surrounding location points except 250m prior (*p*s < 0. 01), however the post hoc pair-wise comparisons showed the effect varied in intensity across the locations.

These post hoc pair-wise comparisons also revealed a condition effect between the Control and Passenger conditions 100m prior to the hazard (p < 0.05), with participants driving significantly faster when in the Passenger condition than during the Control condition at this particular location point. No significant interactions were found at any of the location points measured at Hazard 3 (ps > 0.05). A between-subjects effect was found 100m prior to the hazard ($F_{(1,22)} = 4.504$, $p < 0.05$, $\eta_p^2 = 0.170$), suggesting practiced drivers responded earlier to this hazard, as this would explain why these drivers exhibited significantly slower speeds at this particular location point only.

Reaction time and time-to-collision. As per Experiment 1, reaction times and times-to collision were calculated for each participant based on a pre-determined Reaction Time equals Zero (RT=0) location for each hazard. For Hazard 1, this RT=0 location remained unchanged from Experiment 1, and was set at the location at which the hazard first came into view, 220m prior to where the police car was located. Although the location of Hazard 2 also did not change from Experiment 1 to Experiment 2, the new configuration meant an RT=0 location of 300m prior to the hazard (as opposed to 150m in Experiment 1) was more appropriate. The RT=0 location for Hazard 3 was set at 400m prior to the start of the road works, just as they came into view.

In Experiment 1, only a small percentage of participants registered a deceleration response to the police car. The reduction in speed limit at this hazard in Experiment 2 saw the percentage of participants responding to this hazard increase from 12.5 percent in Experiment 1 to 50 percent in Experiment 2, but despite this increase it was still not feasible to include this hazard in the current analysis. Therefore, shown below are the results for reaction time and time-to-collision at Hazards 2 and 3 only.

At Hazard 2, five participants did not register a reaction to the hazard during one of their trials, but review of the session recordings and simulator output suggested they were travelling at a speed that enabled them to negotiate the hazard safely without reacting. At Hazard 3, a different five participants did not register a reaction to the hazard during one of their trials because they were already reacting to another road feature as they passed the RT=0 location. In each of these cases, the missing data were replaced with the appropriate mean. The participant from the Practiced group who displayed excessive speeds also failed to register a deceleration response at both Hazards 2 and 3 when in the Passenger condition. The procedure set in Experiment 1 determined that a participant who fails to respond to a hazard despite needing to do so should be assigned a maximum reaction time and a time-to-collision of zero. However, as once again it seemed that this could possibly obscure any effects that may be present, this participant was instead assigned the mean reaction time and time-to-collision for the Practiced Passenger condition for both hazards.

Hazard 2. Figure 33 shows the reaction times of participants at Hazard 2, by both conversation condition and group. The Novice group produced what could be considered to be the expected pattern, with the Control condition producing the shortest reaction times; the Passenger condition showing a slight increase from that; the Cell phone condition producing the longest reaction times; and the Videophone condition producing slightly shorter reaction times than the Cell phone condition but not as short as those in the Passenger condition. There appears to be a direct relationship between speed and reaction time for novice participants at this hazard, with the conversation conditions which produced the fastest speeds also producing the faster reaction times and vice versa. This suggests that participants in this group were reacting at the same location prior to the hazard, regardless of conversation condition. If this is true, a faster speed would result in the reaction location being reached faster and would therefore produce a faster reaction time, a trend that can be seen when Figures 29 & 30 and 33 & 34 are compared simultaneously.

The Practiced group, however, produced a very different pattern of results. There was a noticeable difference between reaction times in the Cell phone and Videophone conditions for practiced drivers at Hazard 2, despite these drivers producing similar speeds at this hazard when driving in the same two conversation conditions. Figure 33 shows practiced drivers displayed the longest reaction times in the Control and Videophone condition, while in the Passenger and Cell phone conditions they produced reaction times faster than the Novice group means for every condition.

Figure 34 shows the times-to-collision by conversation condition for both groups at Hazard 2. An almost direct inverse relationship between time-to-collision and reaction time can be seen for novice drivers at this hazard, which provides further evidence to suggest reaction times are a by-product of speed for this particular group of drivers. While the pattern for the Practiced group also shows an inverse relationship, closer examination reveals greater proportionate differences in times-to-collision between the various conversation conditions than can be seen in reaction times for the Practiced group. This again suggests that some factor other than speed was affecting reaction times for this group; however the pattern is not indicative of the expected influence of the types of conversation studied in this experiment. This idea will be explored further in the Discussion.

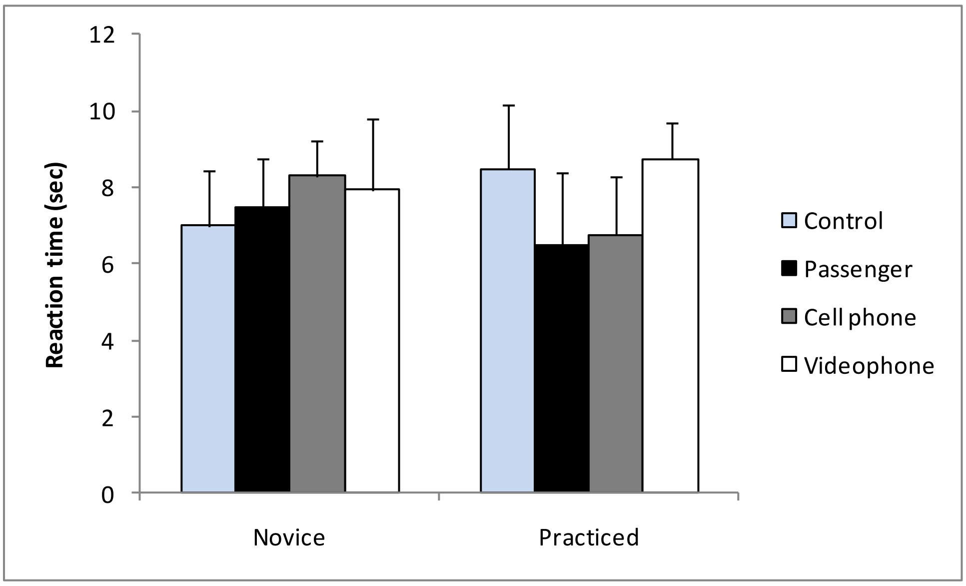


Figure 33. Mean reactions times at Hazard 2, by both conversation condition and group. Error bars show 95% confidence intervals.

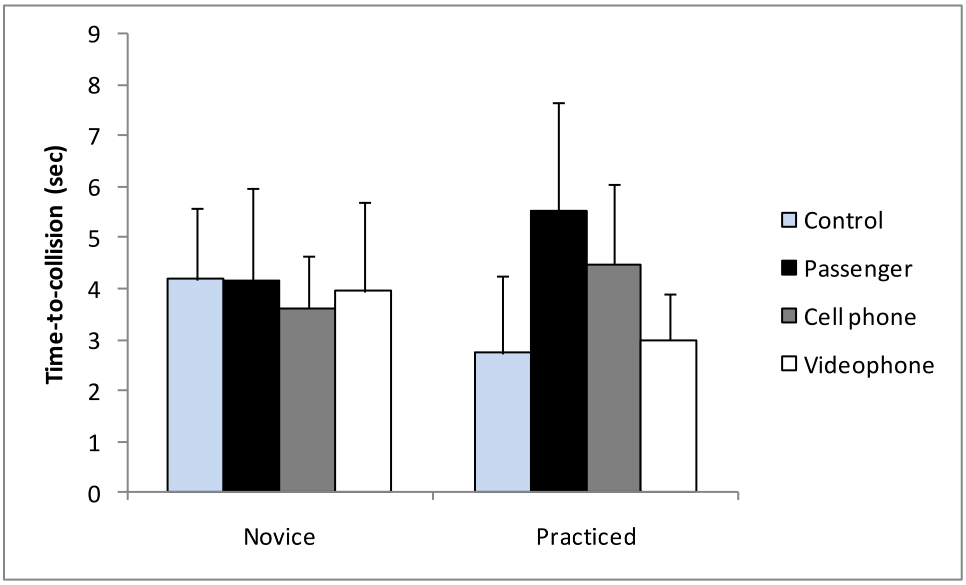


Figure 34. Mean time-to-collision values for Hazard 2, by both conversation condition and group. Error bars show 95% confidence intervals.

As was the case with all driving measures, a 2 x 4 mixed design multivariate analysis variance was used to compare reaction times between the groups and conversation conditions at this hazard. A separate multivariate analysis of variance of identical design was used to compare times-to-collision at this same hazard. The analysis failed to reveal any within- or between-subjects effects for reaction time ($F_{(3,66)}$ = 2.088, p > 0.05, η_p^2 = 0.087; and $F_{(1,22)}$ = 0.009, p > 0.05, η_p^2 = 0.000 respectively). A significant interaction was uncovered as a result of the Control and Videophone means increasing from Novice to Practiced while the Passenger and Cell phone means decreased between the groups ($F_{(3,66)}$ = 3.387, p < 0.05, η_p^2 = 0.133). Again, no significant within- or between-subjects effects emerged in the analysis of times-to-collision ($F_{(3,66)}$ = 2.340, p > 0.05, η_p^2 = 0.096; and $F_{(1,22)}$ = 0.004, p > 0.05, η_p^2 = 0.000 respectively), although the interaction approached significance ($F_{(3,66)}$ = 2.529, p = 0.65, η_p^2 = 0.103).

Hazard 3. Figures 35 and 36 show the reaction times and times-to-collision for participants at Hazard 3 by both group and conversation condition. This time, the Practiced group appears to display the expected trend in reaction times, with the Control and Passenger conditions producing the fastest reaction times and the remote conversation conditions resulting in longer reaction times. There is also a clear practice effect, with the Practiced group showing faster reaction times overall when compared to the Novice group. Within the Novice group, an unexpectedly long mean reaction time for the Passenger condition interrupts what would otherwise be a normal trend. This difference in relative positioning of the Passenger condition between groups causes an interaction, with a negative difference between the Passenger and remote conversation condition means shown in the Novice group, while a positive difference can be seen in the Practiced group. The order of the two remote conditions also reverses between groups.

The between-subjects practice effect is also noticeable in the time-to-collision means, shown in Figure 36. Aside from this observation, the data appears to be normally correlated with the reaction times shown in Figure 35.

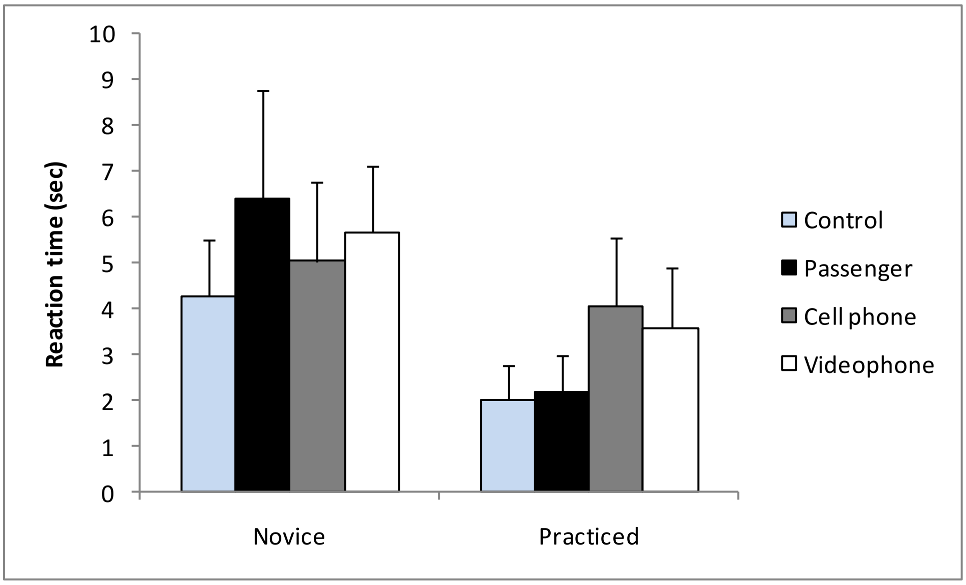


Figure 35. Mean reaction times at Hazard 3, by both conversation condition and group. Error bars show 95% confidence intervals.

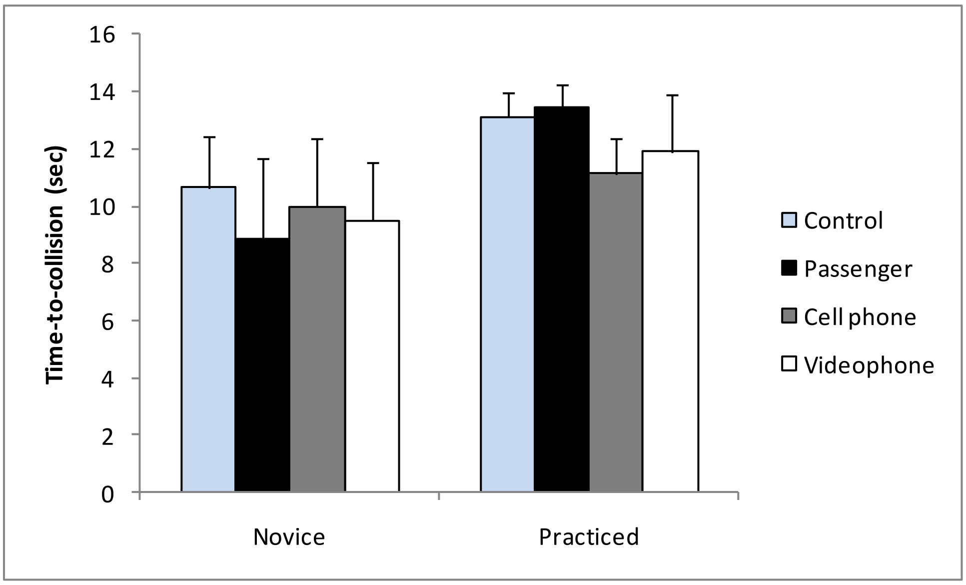


Figure 36. Mean times-to-collision at Hazard 3, by both conversation condition and group. Error bars show 95% confidence intervals.

As before, two 2 x 4 mixed design multivariate analyses of variance were used to compare the various condition-by-group means for reaction time and time-to-collision. The analysis confirmed a significant effect of practice for both reaction time ($F_{(1,22)} = 12.859$, $p < 0.01$, $\eta_p^2 = 0.369$) and time-to-collision ($F_{(1,22)} = 9.797$, $p < 0.01$, $\eta_p^2 = 0.308$). A significant condition effect was also found for reaction time only ($F_{(3,66)} = 3.123$, $p < 0.05$, $\eta_p^2 = 0.124$) and post hoc pair-wise comparisons revealed the source of this effect to be a significant difference between the Control and Videophone conditions ($p < 0.05$), with the use of a videophone while driving resulting in significantly longer reaction times at road works. The relationship between Control and Cell phones also approached significance. The interaction described earlier in this sub-section was also found to be significant for reaction time only ($F_{(3,66)} = 3.006$, $p < 0.05$, $\eta_p^2 = 0.120$).

Crashes. The changes introduced in Experiment 2 saw a dramatic reduction in the number of crashes, to the point they became negligible. Only two crashes occurred in the 96 trials that formed Experiment 2.

Strategic task: rest area. The within-subjects design implemented in Experiment 2 also led to a substantial reduction in the number of people missing the rest area. Only two drivers failed to enter the rest area, with one driver missing the rest area during just one of their four trials and the other during two of their four trials. While this may have been a result of cognitive overload (with the assessment of this being the intended function of this task), the occurrence of this was so low analysis was both unfeasible and unnecessary.

This is in contrast to the findings of Drews et al. (2008), whose experiment showed drivers in the cell phone condition were four times more likely to miss the exit that was the basis of the strategic task in their experiment. It is possible that the failure of the current experiment to reveal any significant differences in this task was due to the within-subjects design.

Conversation performance

The conversation of drivers was assessed using three measures: utterance length, percent of situation awareness utterances (SA utterances) and the total number of words spoken by the dyad during the examined time-window. As per Experiment 1, each of the conversation measures was extracted from transcriptions of the 20 seconds of conversation surrounding each hazard (ten seconds before and after). Similar to the Experiment 1 findings, the occurrence of pauses in conversation was extremely low and as such was excluded as a measure.

Although the hazards were analysed independently for measures of driving performance, it seemed more appropriate in the case of conversation to increase the accuracy of the results by combining the data collected at the three hazards to create an overall set of means. In the case where the conversation data were missing for a whole trial, the missing data were replaced with the appropriate group means. In the case where a crash interfered with the natural time-course of the conversation and a participant was missing conversation data for just one hazard of a particular trial (both crashes occurred just prior to the road works), the missing data was replaced with the mean of that participant's data from Hazards 1 and 2 of the same trial. As per Experiment 1, data for drivers and conversers were analysed independently.

Utterance length. Figures 37 and 38 show the mean utterance length for drivers and conversers by conversation condition and group. The method by which utterance length was calculated was altered slightly for Experiment 2. In Experiment 1, any utterance that at least partly fell within the 20-second window was included in the utterance length calculation, but its length was limited to the number of words that also fell within the window. That is, if a participant made a 15-word utterance but only 5 words formed part of the 20 seconds being examined, a 5-word utterance was recorded. This resulted in the utterance length means being artificially deflated. To rectify this, a new method of dealing with utterances truncated by the 20-second boundaries was introduced in Experiment 2. In Experiment 2, if less than ten words of any given utterance fell within the measured timeframe at a given hazard, the utterance was ignored. If however, ten or more words of an utterance fell within the 20-second window, the utterance was counted in its entirety and the full length was recorded. This resulted in much longer utterance lengths being produced in Experiment 2, but this is likely a more accurate representation of the actual conversation form.

Figure 37, which displays the mean utterance lengths for drivers in each group under the various conversation conditions, highlights what appears to be an interaction between group and conversation condition, with the Videophone condition resulting in longer utterances than the Passenger condition for drivers in the Novice group, but shorter utterances for drivers in the Practiced group. Figure 38 shows a different outcome for conversers, with a group effect being apparent. Conversers in the Practiced group spoke in much shorter utterances than their Novice group counterparts in every conversation condition. It is important to note that the conversers themselves did not differ in level of practice in any way, therefore any group differences in converser data must be a result of changes in driver behaviour.

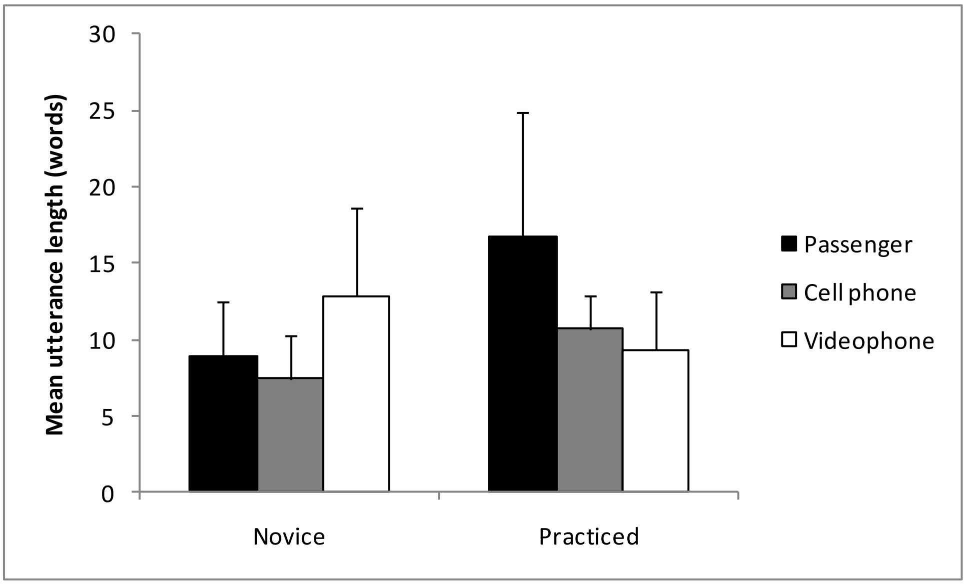


Figure 37. Mean utterance length of drivers by conversation condition and group. Error bars show 95% confidence intervals.

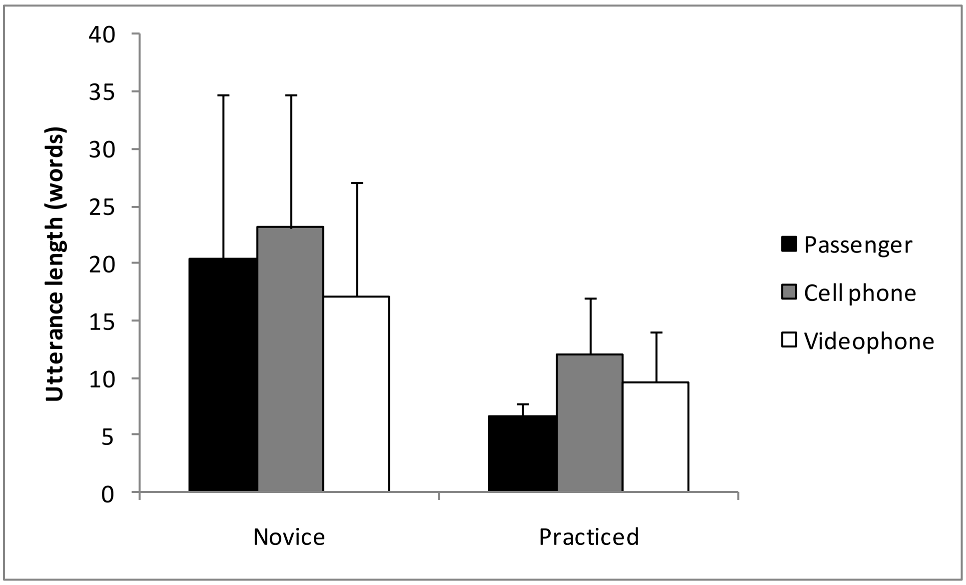


Figure 38. Mean utterance length of conversers by conversation condition and group. Error bars show 95% confidence intervals.

Each of the conversation measures were analysed using a 2 x 3 mixed design multivariate analysis of variance, with two levels of the between-subject factor, practice, and three levels of the within-subject factor, conversation condition. A separate MANOVA was conducted for drivers and conversers for each measure. The analysis confirmed that for utterance length, a significant interaction between level of practice and conversation condition was present for drivers ($F_{(2,44)} = 3.578$, $p < 0.05$, $\eta_p^2 = 0.140$), while a significant practice effect was present for conversers ($F_{(1,22)} = 4.457$, $p < 0.05$, $\eta_p^2 = 0.168$). No other significant effects or interactions were found (ps > 0.05).

Situation awareness utterances. Figures 39 and 40 show the percent of utterances in which the topic of conversation was the immediate driving situation, for drivers and conversers respectively. Both figures show a clear condition effect, with both members of each dyad making a substantially larger number of SA utterances when in the Passenger condition than in the remote conversation conditions. For both drivers and conversers, this effect is slightly reduced by increased levels of experience.

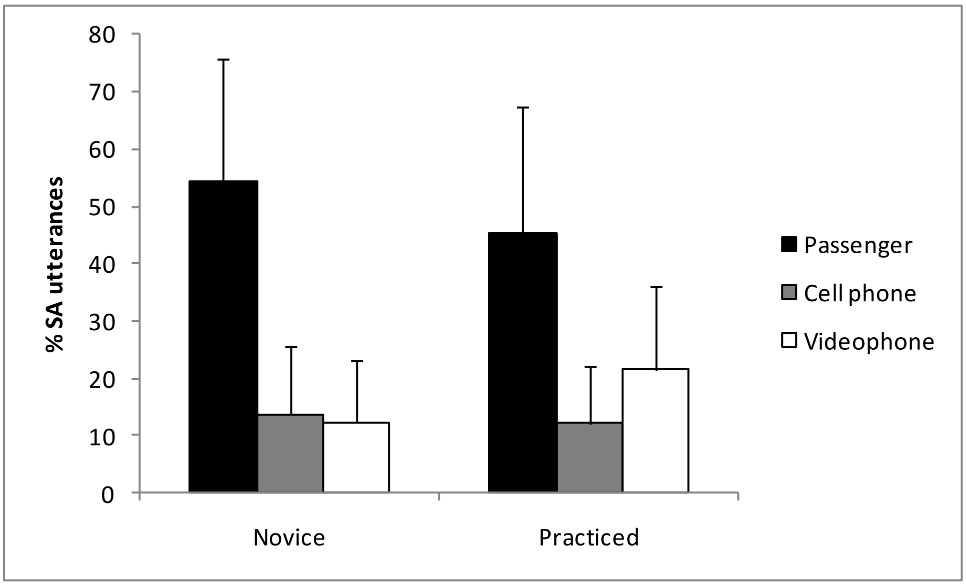


Figure 39. Mean percent of SA utterances made by drivers, by conversation condition and group. Error bars show 95% confidence intervals.

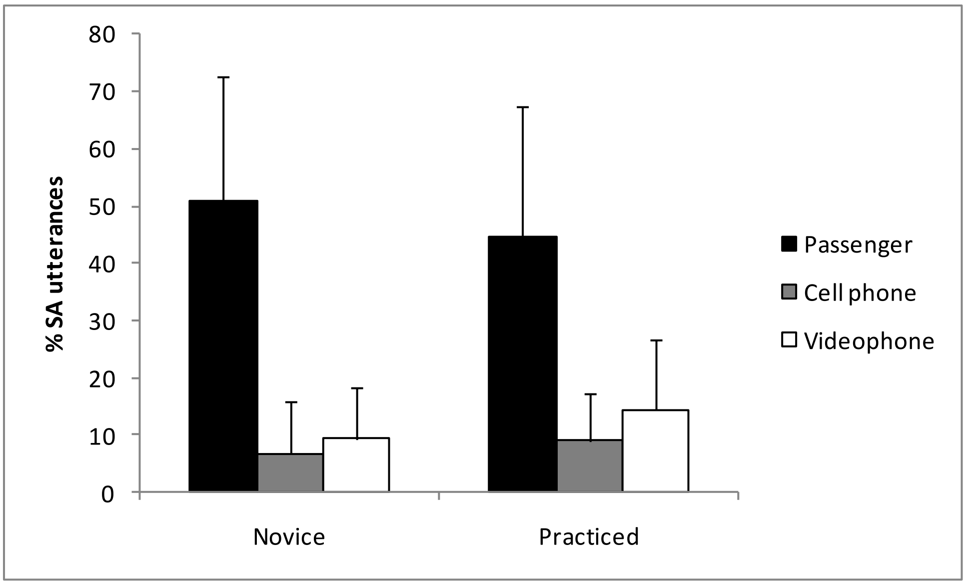


Figure 40. Mean percent of SA utterances made by conversers, by conversation condition and group. Error bars show 95% confidence intervals.

As mentioned previously, a 2 x 3 mixed design multivariate analysis of variance was used to compare the percent of SA utterances across groups and conversation conditions. The results of this analysis confirmed the presence of a significant condition effect for both drivers ($F_{(2,44)}$ = 24.780, p < 0.001, η_p^2 = 0.530) and conversers ($F_{(2,44)}$ = 24.986, p < 0.001, η_p^2 = 0.532). These are the largest effects found across all the measures included in this study. No between-subjects effects or interactions were found for either drivers or conversers (ps > 0.05).

Total number of words. Figure 41 shows the mean total number of words spoken by participant dyads during the 20 seconds surrounding the hazards. The figure shows very similar means across both conversation condition and group. A slight difference in total number of words spoken between groups can be observed, with the Practiced group uttering slightly fewer words than the Novice group. It must be pointed out that the total number of words measure included only those words that fell within the 20-second window, regardless of whether surrounding words had been included in utterance length calculations.

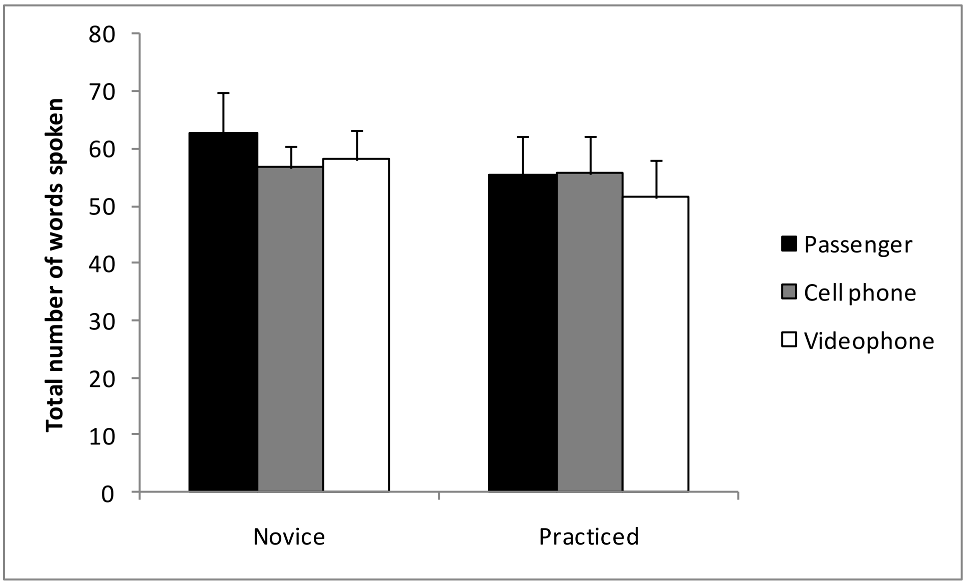


Figure 41. The mean total number of words spoken by participant dyads during the 20 seconds surrounding the hazard, by conversation condition and group. Error bars show 95% confidence intervals.

A single 2 x 3 multivariate analysis of variance was used to compare the total number of words spoken by participant dyads. The results failed to reveal any significant effects (within: $F_{(2,44)} = 2.077$, $p > 0.05$, $\eta_p^2 = 0.086$; between: $F_{(1,22)} = 2.620$, $p > 0.05$, $\eta_p^2 = 0.106$) or any significant interactions ($F_{(2,44)} = 1.223$, $p > 0.05$, $\eta_p^2 = 0.053$).

Ratings

Ratings of conversation difficulty and interference were collected from both drivers and conversers for each of the conversation conditions. Drivers were also asked to provide a rating of driving difficulty at the conclusion of each trial. A number of outlier values appeared in the results, which were not removed from the analysis but were instead highlighted in the figures.

Driving difficulty. Figure 42 shows the ratings of driving difficulty given by drivers at the end of each drive. The figure shows a degree of variability displayed by drivers from both groups across conversation conversations; however both groups rated driving in the Control condition as easier than any of the other conditions.

A 2 x 4 mixed design multivariate analysis of variance was used to compare the mean ratings by both conversation condition and group, and this revealed that the difference between the Control condition and the three conditions involving conversation was indeed significant ($F_{(3,66)} = 8.590$, $p < 0.001$, $\eta_p^2 = 0.281$). The analysis failed to revealed any significant group effects ($F_{(1,22)} = 0.157$, $p > 0.05$, $\eta_p^2 = 0.007$) or interactions ($F_{(3,66)} = 0.469$, $p > 0.05$, $\eta_p^2 = 0.021$).

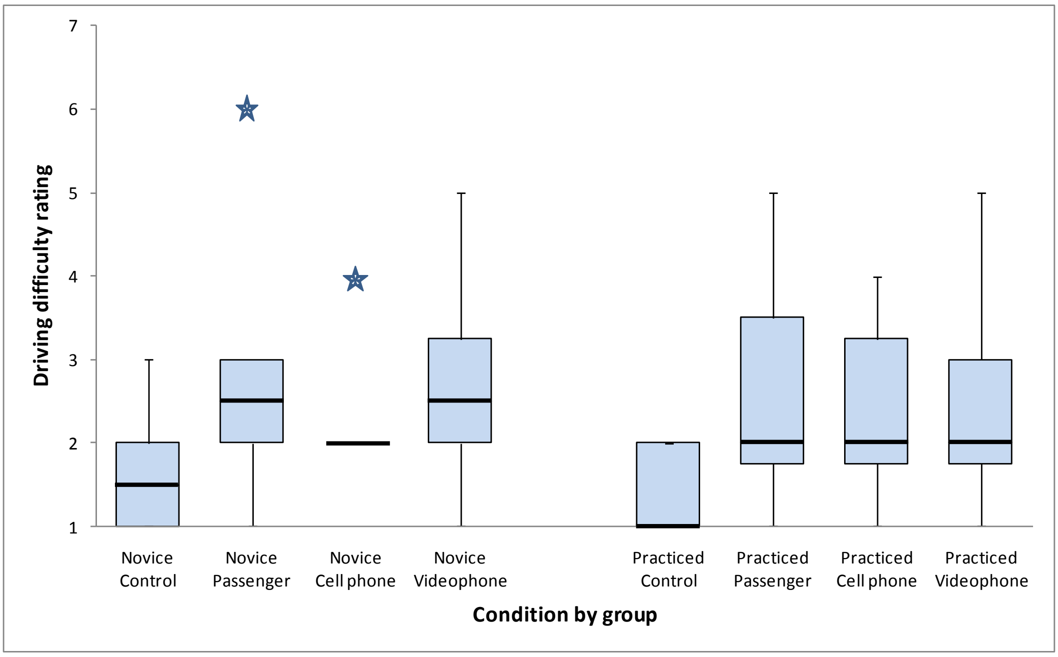
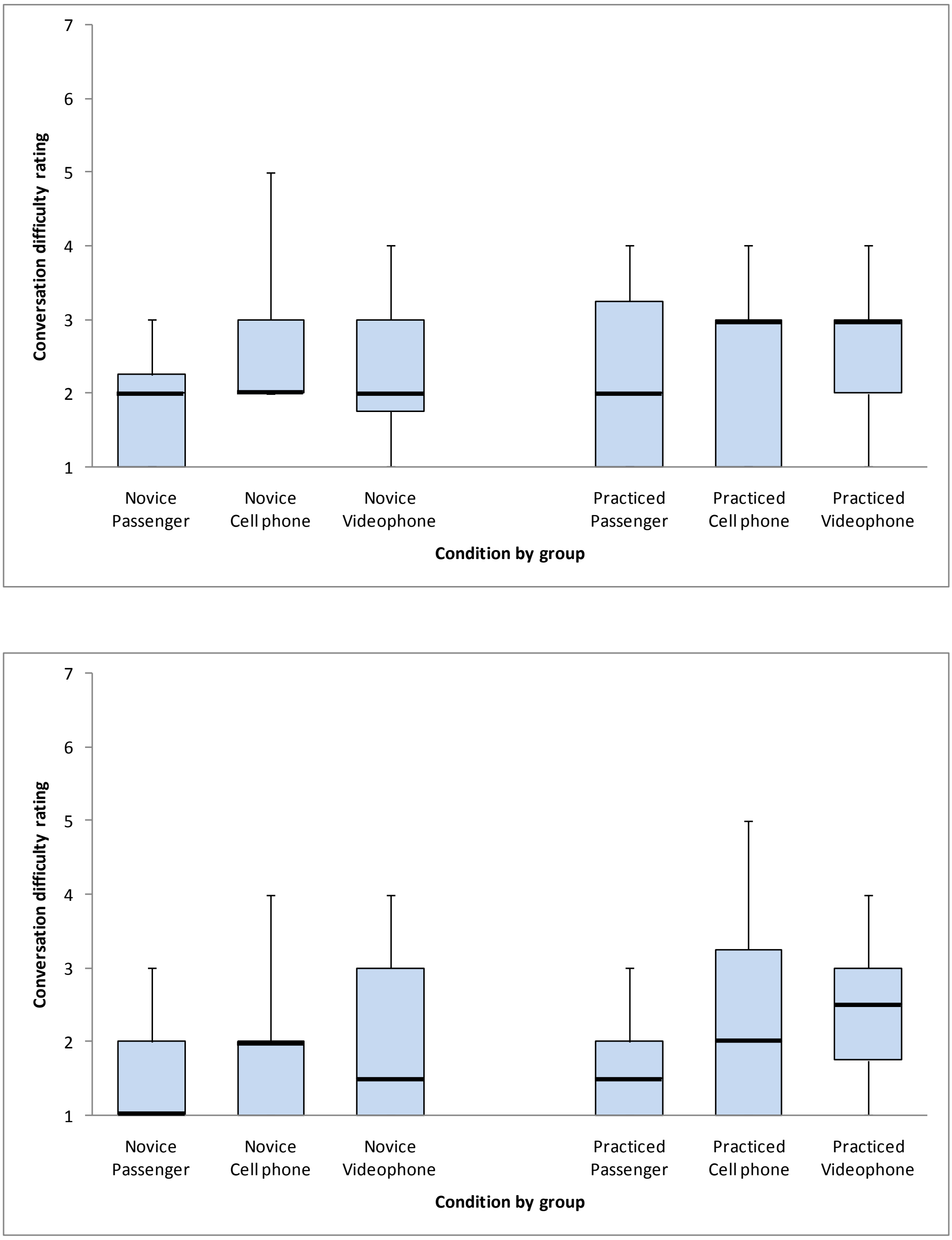


Figure 42. Drivers' ratings of driving difficulty, where 1 is "easy" and 7 is "impossible." Bold lines indicate the median, shaded areas indicate the inter-quartile range, and whiskers indicate the range. Stars indicate outliers, which include any value more than 1.5 times the inter-quartile range greater than the next closest score.

Conversation difficulty. Figure 43 shows ratings of conversation difficulty given by drivers and conversers, by conversation condition and group. The figure shows differing levels of variability across conditions and groups, with no clear pattern emerging for drivers' rating of conversation difficulty. In regards to conversers, Figure 43 indicates novice conversers found the Passenger condition easier than the Videophone, while this pattern was not clear for practiced conversers. The figure also shows practiced conversers showed a greater degree of variability than their novice counterparts in the Passenger and Cell phone conditions.



Figures 43. Drivers' (top panel) and conversers' (bottom panel) ratings of conversation difficulty, where 1 is “easy” and 7 is “impossible.” Bold lines indicate the median, shaded areas indicate the inter-quartile range, and whiskers indicate the range.

A 2 x 3 mixed design multivariate analysis of variance was used to compare the means of drivers' ratings of conversation difficulty, by conversation condition and group. This analysis failed to reveal any significant condition ($F_{(2,44)}$ = 1.305, $p > 0.05$, $\eta_\rho^2 = 0.056$) or group effects ($F_{(1,22)}$ = 0.077, $p > 0.05$, η_ρ^2 = 0.004). It also failed to reveal any significant interactions ($F_{(2,44)}$ = 1.348, $p > 0.05$, $\eta_\rho^2 = 0.058$).

Identical analysis was used to compare converser means, and this time a significant condition effect was revealed ($F_{(2,44)}$ = 4.148, $p < 0.05$, $\eta_\rho^2 = 0.159$), with post hoc pair-wise contrasts made using the Bonferroni adjustment method showing that conversers found talking to the driver as a passenger almost significantly easier than talking to the driver via a videophone ($p = 0.051$). The analysis failed to show any significant group effects ($F_{(1,22)}$ = 1.607, $p > 0.05$, η_ρ^2 = 0.068) or interactions ($F_{(2,44)}$ = 0.106, $p > 0.05$, $\eta_\rho^2 = 0.005$).

Conversation interference. Figure 44 on the following page shows the ratings of conversation interference given by drivers and conversers. Both drivers and conversers in the Novice group showed a smaller range of ratings than their Practiced group counterparts. Also of interest is the fact that no converser in the Novice group provided a rating of 1 ("no interference"), whereas at least one converser in each of the conversation conditions in the Practiced group did provide this rating.

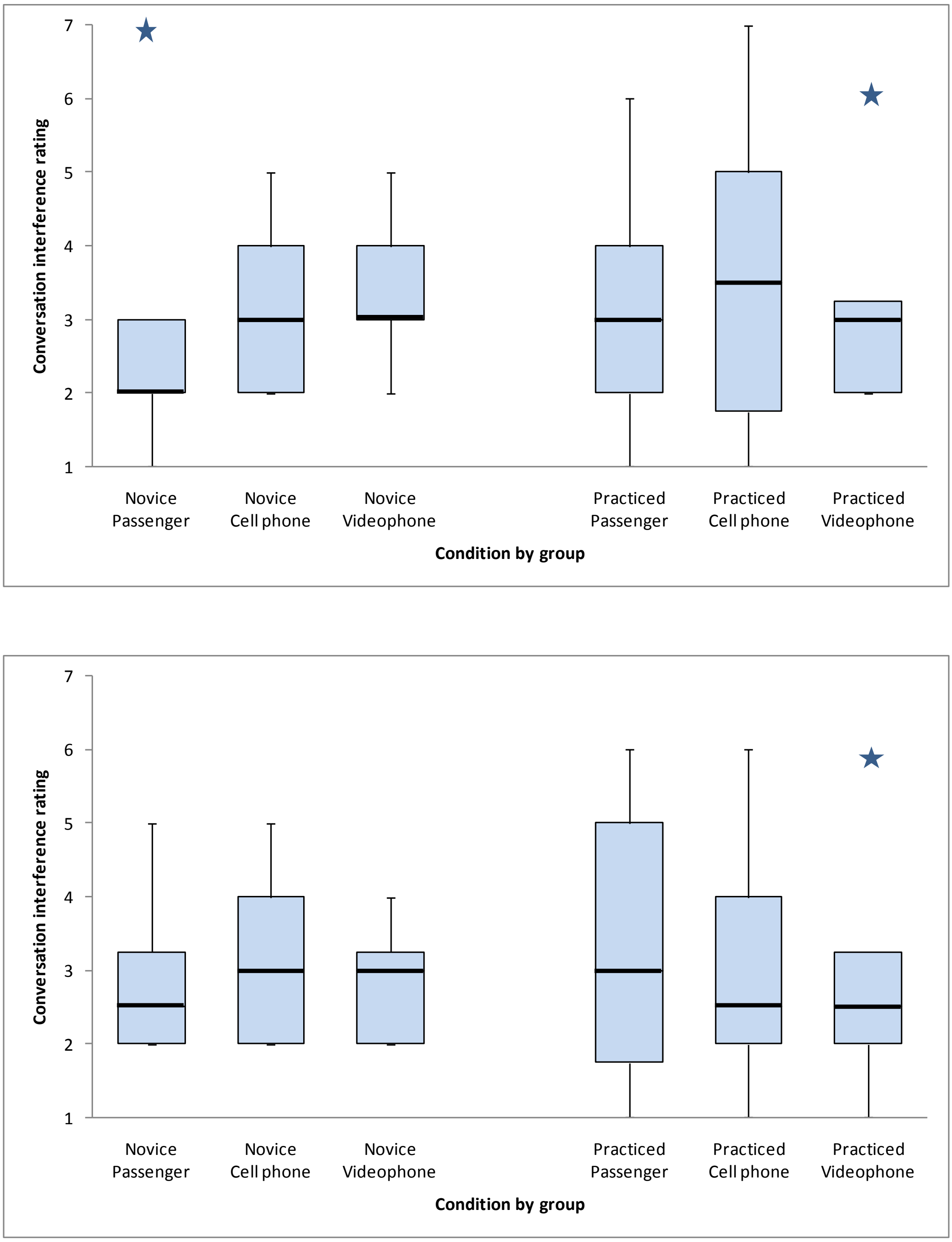


Figure 44. Drivers' (top panel) and conversers' (bottom panel) ratings of conversation interference, where 1 is "no interference" and 7 is "so much interference that driving was impossible". Bold lines indicate the median, shaded areas indicate the inter-quartile range, and whiskers indicate the range. Stars indicate outliers, which include any value more than 1.5 times the inter-quartile range greater than the next closest score.

Again, two 2 x 3 mixed design multivariate analyses of variance were used to compare the means across groups and conversation conditions, with the drivers' and conversers' data being examined in separate analyses. The analyses failed to reveal any significant effects or interactions of any kind for either drivers or conversers (*Drivers* – condition: $F_{(2,44)} = 1.351$, $p > 0.05$, $\eta_p^2 = 0.058$; practice: $F_{(1,22)} = 0.263$, $p > 0.05$, $\eta_p^2 = 0.012$; interaction: $F_{(2,44)} = 1.351$, $p > 0.05$, $\eta_p^2 = 0.058$; *Conversers* – condition: $F_{(2,44)} = 0.084$, $p > 0.05$, $\eta_p^2 = 0.004$; practice: $F_{(1,22)} = 0.005$, $p > 0.05$, $\eta_p^2 = 0.000$; interaction: $F_{(2,44)} = 0.299$, $p > 0.05$, $\eta_p^2 = 0.013$).

Discussion – Experiment 2

The overall effects of videophones

The aim of this research project was to empirically investigate whether a remote conversation that provided drivers and conversers with access to visual cues would result in improved driving performance when compared to a traditional remote conversation (i.e. a videophone conversation versus a cell phone conversation). The changes introduced in Experiment 2 resulted in a well-controlled experiment, more powerful than Experiment 1 given the within-subjects design. However, the findings show that providing drivers and conversers with access to these visual cues did little to alter the effects of remote conversations on driving performance, with the Cell phone and Videophone conditions producing similar results in all of the driving performance measures examined in this study. Conversation measures were also analysed as a means of identifying possible reasons for any observed differences between effects. However, the results show that access to visual information also had little influence on conversation form and content for both drivers and conversers, with the Videophone condition producing results similar to those found under the Cell phone condition for all conversation performance measures.

The failure of this experiment to find any significant differences between the effects of cell phones and videophones does not necessarily mean that the availability of visual cues during passenger conversations does not contribute to minimising the distracting effects of these conversations. Rather, it merely suggests that the addition of visual cues alone is insufficient to replicate in a remote conversation the lower level of distraction found in passenger conversations, or that a videophone is an unsuitable means of supplying access to these visual cues. As hypothesised at the outset, it may be that the addition of visual cues did aid in reducing cognitive load for drivers during the Videophone condition but this effect was cancelled out by the increased distraction caused by the videophone screen. Alternatively, it may be that the videophone served to reduce cognitive load in some drivers while increasing it in others. This second idea is supported by drivers' ratings of difficulty and interference.

In Experiment 2, several participants reported experiencing much lower workload (shown by lower ratings of driving and conversation difficulty and interference) when driving and conversing via videophone compared to using the cell phone; however an equal number of participants found that use of a videophone led to a higher level of workload, while the remaining participants experienced no difference between the two. This combination resulted in no significant differences between the Cell phone and Videophone conditions being found in any of the ratings collected in this study. Many of the drivers who found that the videophone in fact led to reduced cognitive workload reported being surprised by this experience, suggesting that the observed directional differences in perceived workload between participants were not a result of prior expectations, for these participants at least.

Another explanation could be that passenger conversations have a specific combination of factors that work together to minimise distraction (including the converser's ability to see the road, the availability of visual cues and other previously untested ideas such as psychological closeness and conversers' feelings of safety, among others), and these factors must all be present for their effects to be evident. If this were the case, any experiment examining just one of these factors while controlling for the others (such as the current study and Charlton's 2009 experiment) should fail to see any noticeable effects, whereas an experiment that tried to replicate as many of these features as possible in a remote conversation may in fact succeed. Following this line of argument, a videophone conversation that provided conversers with not only a view of the driver's face but also information about the real-time driving situation as it unfolded may produce different results from those observed in this experiment.

Conversation modulation in videophone conversations

One of the theoretical premises upon which this research was based was that conversation modulation is the key distinguishing feature between passenger and cell phone conversations, and that all of the above-mentioned features of passenger conversation serve to evoke this conversation modulation. That is, this experiment was not designed to test whether access to visual cues had a similar or greater effect than conversation modulation on driving performance, but rather whether providing access to visual cues was sufficient to evoke a greater level of conversation modulation than that seen in cell phone conversations.

Drews et al. (2008) showed that drivers modulate their conversation as driving difficulty increases regardless of conversation mode, whereas conversers displayed differing levels of modulation depending on conversation type. Drews et al. suggested this was due to cell phone conversers either failing to recognise and respond to verbal cues given by drivers or failing to respond to them to an appropriate degree. Therefore, it was hypothesised that giving conversers access to visual cues such as changes in drivers' expression or glance rate through the use of a videophone would enable conversers to recognise and respond to verbal cues more readily, resulting in greater levels of conversation modulation on the part of the conversers. However, during Experiment 2, driving participants were told they were free to look at the screen as much or as little as they wished, which resulted in the majority of participants choosing not to look at the screen at all while driving. This seeming lack of attention paid by drivers to the conversation may have given conversers the opposite impression to that predicted, leading them to believe the conversation was imposing even less interference on the driving task than the verbal cues alone may have suggested, and suggesting there was in fact less need to alter their conversation as opposed to more. This idea is reflected in conversers' ratings of conversation interference.

While converser ratings of interference were similar across conversation conditions, conversers did rate the Videophone condition as the least interfering of all the conversation conditions, with the Passenger condition being rated as the highest. While this difference never reached statistical significance, it could possibly be interpreted as further evidence to support the idea that visual cues alone are insufficient for conversers to form an accurate impression of conversation interference, with information about driving performance being a necessary component, as suggested earlier.

It is also interesting to note that the results of this experiment did not line up with Drews et al.'s (2008) findings regarding drivers, with drivers in this experiment also showing reduced levels of conversation modulation in the remote conditions, particularly in terms of situation awareness utterances. SA utterances were the one of the main measures of modulation used by Drews et al., and it may be that coding differences led to the discrepancy between the current study and their work. Drews et al. counted the number of initial references to the driving scenario as well as the number of subsequent turn-takes made by driver and converser in response to the initial traffic-related comment. On the other hand, this experiment measured SA utterances as a percentage of total utterances; therefore a conversation resulting from a single SA utterance made by the driver would result in a high percentage of SA utterances for both parties, whereas a situation in which the driver made an SA utterance but the converser did not respond would result in the driver having a low percentage of SA utterances and the converser having a percentage of zero (assuming the driver did not enter into a monologue of utterances about the driving situation). Thus, while both of the cases described above would have resulted in just one initiating utterance being recorded for the driver in Drews et al.'s study, the driver would have obtained very different percentage scores in the current study.

Additional findings of importance

Despite failing to reveal any significant differences between the effects of cell phone and videophone conversations in any of the driving or conversation performance measures, the results of Experiment 2 still have important implications for understanding the distracting effects of cell phones. These implications extend beyond the videophone-related findings discussed above, with the improved design of Experiment 2 allowing for the observation of many additional unexpected but potentially useful findings.

In order to adequately describe a series of unexpected results, it is first necessary to reiterate what was expected. What was not stated explicitly in the Introduction, but was definitely implied, was the expectation that regardless of the outcome of the Videophone condition, the relationship between the remaining three conversation conditions – Control, Passenger and Cell phone – would resemble those found by other experiments adopting similar methodologies. Based on the outcomes of previous research into the relationship between passenger and cell phone conversations, it was proposed that experiments with high ecological validity were able to detect differences between the effects of passenger and cell phone conversations, while experiments with lower ecological validity were not. As every effort was made to ensure this experiment was as ecologically valid as possible, it was expected that this experiment would also produce results showing a distinct difference between cell phone and passenger conversations, with passenger conversations resulting in driving performance that resembled that produced under the Control condition.

Although speed as a measure of driving performance has produced mixed results in the past, with some research suggesting drivers increase their speed while conversing on a cell phone and other research suggesting cell phone use leads to a reduction in speed, it was expected that this particular study would see an increase in speed when drivers were conversing on a cell phone compared to driving in the Passenger and no-conversation Control conditions. This expectation was based upon the fact that this was what was found by Charlton in 2009, and the current study was modelled very closely on Charlton's research, to the point that the same simulator, road, and hazards were used and a very similar experimental protocol was followed.

There were some differences between the two studies however, which naturally resulted in slightly different expectations. Two major differences between the current study and Charlton's experiment (2009) were the switch to within-subjects design and the inclusion of a sample of practiced drivers. It was thought that the introduction of the within-subjects design might result in a reduction in the size of any observed effects in driving performance, with the added possibility that any condition effects existing in reaction time could be completely obscured due to drivers being able to anticipate the upcoming hazards in three of the four trials they completed. If reaction time effects were not obscured, it was thought that drivers would react significantly faster in the Control condition than in the Cell phone condition, with reaction times in the Passenger condition resembling those of the Control condition. In regards to the Practiced versus Novice groups, it was expected that practiced drivers would display greater competence in vehicle handling, which would be useful for obtaining usable data void of crash interference should the changes to the simulation scenario be insufficient for reducing the crash rate to an acceptable level. Aside from this, it was expected that both groups would produce similar patterns of driving and conversation behaviour.

The pattern of results obtained at Hazard 3 (the road works) was most in line with this set of expectations. At this hazard, drivers from both groups drove significantly faster in the remote conversation conditions than in the Control condition. They also reacted significantly faster in the Control condition than in the Videophone condition, with this effect almost reaching significance between the Control and Cell phone conditions also. The Passenger condition revealed some unexpected findings however, with novice drivers following the expected pattern and driving at speeds similar to the Control condition when driving with a passenger; whereas practiced drivers exhibited the fastest speeds of any condition when driving with a passenger. An unexpected practice effect was also revealed in reaction times at this hazard, with drivers in the Practiced group responding much faster to the hazard than novice drivers, regardless of condition. These two unexpected findings will be discussed individually, beginning with the reaction time finding.

Reaction times, slowed speech, and cognitive workload. The significant practice effect observed in mean reaction times may be partially explained by the presence of a corresponding practice effect in the total number of words spoken at the same hazard. Practiced dyads uttered fewer words in the 20 seconds surrounding Hazard 3 than did their Novice group counterparts. This measure was included to represent both a slowing of speech and an increase in the number of short pauses, be they mutual or initiated by one party. This coupling of results suggests that a reduced number of words is related to shortened reaction times, but it does not shed light on the direction of this relationship. That is, it may be that drivers responded quickly and this was then followed by a mutual slowing of speech rate, or it may have been that the speech pattern was adjusted first, which freed up drivers' cognitive resources and enabled them to respond faster.

While a relationship between changes in speech rate and faster reaction times was in-and-of-itself not unexpected but in fact predicted, phenomena such as this were not expected to be dependent on level of practice, especially considering members of the Practiced group were no more practiced at driving and conversing than members of the Novice group. (The previous experiment from which the Practiced group drivers were recruited was not a conversation-based experiment). This leads to a third possible explanation; rather than a change in one of these factors leading to a change in the other, it may be that both of these factors are related to a separate third factor and the observed changes in driving and conversation behaviour occurred simultaneously as a result of a change in this third variable.

It is possible that this third variable is cognitive workload. Authors in support of the conversation modulation hypothesis have suggested that the very reason drivers and conversers alter conversation is to reduce the cognitive load experienced by the driver, making way for safer driving (e.g., Charlton, 2009, Drews et al., 2008). Yet this does not explain why only drivers in the Practiced group reacted faster and changed their conversation at this hazard. Given the increased experience of drivers in this group, it is reasonable to assume that at any given time, practiced drivers were experiencing a *lower* level of cognitive workload induced by the driving task than their Novice group counterparts. It would therefore stand to reason that if these changes were a mere reaction to experienced levels of workload, this effect should have been more pronounced in novice drivers than practiced.

This suggests that it was not changes in the current level of perceived workload that led to these particular changes in driving and conversation behaviour, but rather it was the practiced drivers' ability to accurately *anticipate* upcoming changes in workload and determine the appropriate response that resulted in the significant differences observed at Hazard 3. Research has shown that less-experienced drivers are more likely to overestimate their capabilities and underestimate the objective demand of driving tasks (e.g., Deery, 1999; Gregersen, 1996). The combination of these two factors may have resulted in the novice drivers underestimating the need to adjust their behaviour to negotiate the upcoming set of road works.

The notion that drivers calculate upcoming shifts in workload and make compensatory behaviour changes based on these calculations suggests drivers are in fact seeking to maintain a target level of cognitive workload – an idea that was first introduced in the Task-Capability Interface (TCI) model of driver decision-making (see Fuller & Santos, 2002, or Fuller, 2005, for a complete explanation of the model). At the centre of this model is the concept of task difficulty homeostasis – the notion that drivers seek to maintain task difficulty within a target range, and make compensatory changes any time they anticipate or experience task difficulty breaching the range boundaries. The concept of task difficulty used in this model is not an objective rating of the driving scenario, but rather is said to be the product of the driver's capabilities and objective task demand (or the driver's perception of both). That is, two drivers with differing capabilities would assign different levels of task difficulty to the same hazard. The same two drivers would also be expected to experience differing levels of cognitive workload in response to the same hazard, suggesting the concepts of task difficulty (as used in this case) and cognitive workload are interchangeable, with Fuller himself acknowledging the similarity of these two concepts in his 2005 article. Fuller and Santos' model also states that the boundaries between which the desired level of workload lies are also a product of driver experience and capability, which helps to explain the differences between novice and practiced drivers' behaviour at Hazard 3. That is, the upper limit of acceptable workload was likely to be lower for practiced drivers, based on a more realistic view of the task demands and their own capabilities.

The TCI model (Fuller and Santos, 2002) has several limitations that have implications for its application to these findings. The first is that the model gives very little attention to the role of secondary tasks, instead focusing mainly on driving behaviour (speed in particular). However, it is possible for a driver to reduce overall cognitive workload without altering driving behaviour by reducing the workload induced by a secondary task (e.g., through conversation modulation, or turning down the radio while searching for a particular street). Alternatively, if performance on the secondary task was of greater importance, drivers may choose to reduce overall workload by devoting fewer resources to certain aspects of the driving task, such as speed maintenance. This is turn could result in an overall increase or decrease in speed, depending on the road design or individual characteristics of the driver.

Fuller and Santos (2002) do not acknowledge that in certain circumstances maintaining a slower speed may in fact increase workload but instead suggest that the relationship between speed and workload is unidirectional – that is, as speed increases, workload (or task difficulty) increases. They even go so far as to say that if task difficulty is too low, drivers will increase their speed. However, a more logical explanation than drivers artificially inflating the difficulty of the driving task until desired workload is reached could be that drivers may instead choose to increase workload by adding a secondary task – such as engaging in a conversation or listening to music or an audio book.

While the focus of the above explanation is on anticipated changes in workload, it is not to say that actual changes in workload do not also influence driving or conversation behaviour. In fact, it may be that when drivers are conversing on a cell phone their ability to anticipate changes in workload is impaired (perhaps due to an impaired ability to process upcoming changes in the driving scenario if inattention blindness is at work – see Strayer et al. (2003) for a more detailed explanation of this concept); therefore any compensatory behaviour changes made by these drivers are in response to changes in real-time workload experience. These changes would likely occur very close to the hazard itself, but may still be sufficient to allow for safe negotiation. This would explain why almost all drivers managed to negotiate the road works successfully in every condition despite displaying slower reaction times and faster speeds in the Cell phone and Videophone trials. Novice drivers may have been even more reliant on actual changes in workload in the remote conversation conditions due to their discrepancies between actual and anticipated workload. (Fuller also suggests a more reactive explanation of behaviour for newer drivers in his 2005 piece).

The relationship between speed and utterance length. Conversational measures may also be associated with the second unexpected finding to be observed at Hazard 3, which was the unusual pattern of speed behaviour displayed by practiced drivers in the Passenger condition. As mentioned earlier, the practiced drivers' speeds under the Passenger condition were at the opposite end of the range from the Control condition, whereas the speeds of the Novice group in the Passenger condition resembled the Control condition closely. Further analysis of conversation performance at this hazard revealed that the mean utterance length of practiced drivers was longest in the Passenger condition, and this mean was more than double that of their conversation partners and that of the novice drivers for the same condition. The novice conversers also produced their longest utterances in the Passenger condition at Hazard 3, with their mean utterance length in the Passenger condition being even higher than that of the practiced drivers. Therefore, when approaching and passing through this particular hazard, novice drivers were listening to long utterances in the Passenger condition while practiced drivers were producing them. The corresponding patterns of speed suggest that, contrary to what some researchers have suggested in the past, listening to long utterances may not have a noticeable effect on speed (perhaps if the driver is able to tune them out), while producing them does.

Something that is not explained by the results is the reason why practiced drivers produced long utterances in the Passenger condition only. One possible explanation is that these drivers may have been more excited about the opportunity to share the simulator experience with a friend than novice drivers, especially considering that the practiced drivers had driven alone in the simulator on twenty previous occasions as part of another experiment. They may have even told their current conversation partner about the simulator in the past. Applying the previously-outlined model to this situation, it may have been that during the Passenger trials, practiced drivers were willing to sacrifice certain aspects of driving performance to maintain the conversation, while still maintaining a level of cognitive workload that was beneath their maximum threshold.

It is interesting to note that the effect of utterance length seems to be limited to speed only. Despite producing the poorest speed performance in the Passenger condition, practiced drivers displayed faster reaction times in this condition than in either of the two remote conversation conditions, while novice drivers produced longer reaction times in the Passenger condition than in the two remote conversation conditions. Therefore, this could suggest that different aspects of conversation modulation affect different aspects of driving performance. Alternatively, it may be that in this particular case, the longer utterance lengths are an indication of who was leading the conversation at the time, with practiced drivers being able to place a necessary pause in between two long utterances if they were in fact leading the conversation, while novice drivers were perhaps denied this opportunity if the conversation was being paced by their partner. It is also important to consider that given the coding rules used in this experiment, it is possible that the practiced drivers may have halted their conversation entirely as they passed through the hazard, despite producing long utterances either side of the hazard in the Passenger condition (so long as at least ten words of each of these utterances fell within the 20-second window).

The influence of hazard type on speed. A brief look at the results at Hazard 2 reveals other unexpected findings that may also be relevant to the effects of cell phones on driving. The most obvious difference between the results of Hazard 2 and 3 is that the condition effect found in speed behaviour changes direction. That is, at Hazard 3, the Control condition resulted in speeds that were significantly slower than the remote conversation conditions; whereas at Hazard 2 both practiced and novice drivers drove significantly faster in the Control condition than in the remote conversation conditions. This pattern of results suggests that rather than concurrent cell phone use consistently causing drivers to produce faster speeds as Charlton (2009) found, it may be that concurrent cell phone use leads drivers to produce an overestimation or underestimation of the correct driving response, depending on the nature of the driving situation or hazard.

This explanation still leaves two unanswered questions: 1) what is it about cell phones that cause this response? and 2) what aspect of the driving situation determines whether an overestimation or underestimation of the correct driving response is made? The first question merely reiterates the basis of not only this study but almost all research into cell phones and driving and will not be discussed in any further detail here. The second question cannot be answered with any certainty without further investigation, but there are several possible explanations.

The first is that the two hazards may have differed in their level of difficulty (objective task demand as opposed to task difficulty as used by Fuller and Santos, 2002). If it were to be assumed that Hazard 2 and Hazard 3 differed in their degree of driving difficulty, with Hazard 2 being slightly easier, it could potentially be concluded that concurrent cell phone (or videophone) use leads to a reduction in speed during difficult driving tasks and an increase in speed during easier driving tasks. Conversely, it may have been that Hazard 2 was more demanding and in fact cell phone use leads to an increase in speed during difficult tasks and a decrease during easier tasks.

Unfortunately, as driving difficulty was assessed over the whole trial as opposed to at each hazard, the direction of any potential relationship between speed and driving difficulty cannot be determined, nor can it be known whether participants even considered the two hazards to have differing levels of difficulty. The change in direction of speed results may instead be related to some other feature of the hazards, such as the fact that Hazard 2 presented the chance of colliding with another moving vehicle whereas Hazard 3 included only static obstacles; or perhaps that Hazard 3 incorporated a change in speed limit whereas Hazard 2 did not.

Regardless of the underlying reasons for the above phenomenon, it is interesting to note that the unusual pattern of speed behaviour produced by practiced drivers in the Passenger condition was also present at Hazard 2, despite the direction of the relationship between the speeds of the Control and the remote conversation conditions changing. As was the case at Hazard 3, the speeds of novice drivers in the Passenger condition at Hazard 2 were most similar to the Control condition, whereas the speeds produced by practiced drivers in the Passenger condition were most different from the Control condition. However, as the direction of the overall trend had shifted, this meant that practiced drivers drove at their slowest in the Passenger condition at Hazard 2 while Novice drivers produced some of their fastest speeds in the Passenger condition. The same pattern of utterance length behaviour was also present at Hazard 2, with practiced drivers and novice conversers speaking in longer utterances. The fact that this behaviour persisted across hazards, with the direction of the observed effects also changing as the direction of the relationship between the other conditions changed, provides further evidence for the idea that the effects of any task that impairs performance or alters workload are in fact moderated by the nature of the driving task chosen. Together, these results suggest that the opposing effects of cell phone conversations on speed observed by other researchers may have been a result of the type of driving task chosen as opposed to other methodological differences.

The story told by a lack of results. One final interesting but unexpected finding can be observed in the driving and conversation results from Hazard 1. Surprisingly, no significant differences were found in any of the driving performance measures at this particular hazard despite both condition and practice effects being found in conversers' conversation performance measures. In fact, there were more significant differences in conversation measures at this hazard than at either of the other two hazards. This finding may serve to further extend the idea of driving difficulty moderating the effects of conversation and changes in secondary workload. If it was to be assumed that Hazard 1 was the easiest of the three hazards, with participants being able to negotiate their way through the gap between the police car and truck without necessarily needing to make a deceleration response (shown by the number of participants who chose not to remove their foot from the accelerator as they passed this hazard), it could be said that this particular hazard was too easy to induce noticeable differences in driving performance and any effects of conversation mode were obscured. This in turn would suggest that any attempts to compare the effects of cell phones and passenger conversations on driving performance would fail to observe an existing effect if the driving task used in the experiment was too easy (despite being more complex than driving down a straight road). Thus, even if experimenters were to vary the degree of driving difficulty throughout the experiment, it may be that no effects would be seen unless the 'hard' driving tasks passed a certain level of difficulty, or imposed a certain level of workload. This may explain the lack of findings observed by some researchers investigating the differences between passenger and cell phone conversations.

This may also be indicative of some people's real-world experiences of using a cell phone while driving. In contrast to theories proposed by other researchers that suggest a mis-calibration between actual and perceived levels of driving performance could be the reason some drivers deny cell phone conversation results in driving impairment (Horrey et al., 2008), it may instead be that these drivers have never yet been talking on a cell phone when experiencing a driving situation difficult enough to produce noticeable decrements in their driving. It could even be possible that with increased use of cell phones, drivers have in fact learned to manage the detrimental effects of cell phones, but only in driving situations that are a below a certain level of driving difficulty. It would therefore follow that even if knowledge about the negative effects of cell phone use on driving performance were to be presented to these drivers, it may be that rather than leading to reduced cell phone use on their part, this knowledge may instead serve to create or reinforce a belief that cell phones may negatively affect the driving performance of *other* drivers, but they personally have learnt to drive safely while using a cell phone and are therefore immune to the effects. What these drivers may fail to realise is that should they find themselves talking on a cell phone whilst encountering a driving situation serious enough to induce a readily noticeable level of driving impairment arising from that very cell phone use, their driving skills may not be sufficient.

Alternatively, the reason no significant effects were observed in driving performance at Hazard 1 may have been because drivers did not accept their conversation partner's attempts at conversation modulation. This idea of initiation and acceptance can also be linked to the previously-mentioned theory regarding anticipated changes in driving-induced workload. If upon seeing a hazard in the distance, drivers and conversers came to different conclusions about whether or not this hazard would cause workload to exceed the maximum desired level, the person believing it would exceed the maximum (i.e. the converser in this case) would likely initiate conversation modulation; while the person believing no change needed to be made may override this attempt at conversation modulation by continuing the conversation themselves. For example, if a converser paused mid-sentence in response to an upcoming hazard that the driver anticipated could be negotiated safely without a reduction in workload, the driver may either complete the sentence themselves or ask a question which forced the converser to complete their sentence. Further review of the audio-visual recordings from Experiment 2 revealed a number of these occurrences throughout the experiment.

Summary

Despite the use of videophones providing few clues as to what it is about cell phone conversations that make them so distracting, this research project as a whole has led to the introduction of a number of new potential ideas and theories. These include the notion that the nature of the driving task may affect the direction of the effects of cell phones on driving behaviour – with cell phones potentially leading to an overestimation or underestimation of the correcting driving response depending on the nature of the driving task. Also suggested was the idea that conversation modulation must be initiated and accepted to produce observable effects on driving performance; that objective driving difficulty must be above a certain level for effects of secondary workload to be observed; that features of a passenger conversation may have no observable effects when examined in isolation; that different aspects of conversation behaviour may affect different aspects of driving performance, and lastly, in relation to a specific aspect of conversation behaviour, that listening to long utterances may not affect driving performance while producing them does. The idea of cognitive workload playing an instrumental role in driver performance was also explored, with Fuller and Santos' (2002) Task-Capability Interface model and the notion of task difficulty homeostasis helping to explain some of the findings.

Limitations

As with any research, there are a number of factors that limit the extent to which these results can be generalised. First and foremost is that this research was conducted in a simulator as opposed to using an actual vehicle on real roads. One of the main disadvantages of using of a simulator is the increased likelihood of participants producing driving behaviour that is not an accurate reflection of their real-world behaviour. These changes in behaviour may be intentional, as may be the case if participants fail to treat the simulation as real; or unintentional, if the lack of vestibular feedback affects a participant's vehicle handling, speed control, or acceleration or deceleration behaviours.

However, there are also several advantages to using a simulator. One of the major advantages is the ability to tailor the driving scenario to test a specific hypothesis. Simulators also afford greater control of experimental conditions, with oncoming traffic, lighting, and road surface conditions all able to be kept constant between participants. Liability and safety concerns are also eliminated, as is the ethical issue of needing to obtain informed consent from all other road users who become inadvertently involved in an on-road field experiment.

Another limitation that also may have impacted these results was that the conversation observed in this experiment was essentially forced. Even within studies such as this one that utilise naturalistic conversation, there is always the possibility that being asked to maintain a conversation for the duration of an experimental drive may cause some participants to produce conversation behaviour that does not reflect their normal conversational style. This may have been particularly applicable in this case, given that participants had to complete three trials involving conversation within one session.

Another limitation specific to this study may have been caused by the within-subjects design, which saw participants driving the same road four times consecutively, encountering the same hazards each time. While the resulting familiarity no doubt affected driving behaviour and may be a reason for some of the differences in observed findings between the current research and Charlton's (2009) findings, the move to a within-subjects design may also have increased ecological validity, if one considers that most drivers in the real world spend a greater proportion of time driving familiar roads (such as their daily route to work) than unfamiliar roads, and many of the hazards on these roads are in fact fixed (e.g., a tight corner).

The inclusion of a group of practiced drivers may also aid in explaining the presence of differences in results between this study and Charlton's (2009) work. It could be said that testing participants in an unfamiliar simulator (or car) on an unfamiliar road may possibly result in the overestimation of the effects of cell phones if the unfamiliarity of the vehicle and road produced increased levels of cognitive workload. However, it is also likely that the unfamiliarity may produce in participants a heightened sense of vigilance, which in turn may balance out the effects of the unfamiliarity. This combination may therefore possibly result in relatively accurate results, but which may actually be based on parameters that do not necessarily represent real-world driving. The difference in simulator experience between participants in the current study could be likened to the difference in driving experience found between new and experienced drivers in the real-world. While there is no way of determining whether the hypothesised increased workload and heightened vigilance were actual factors in Charlton's (2009) experiment, or contributed to some of the differences between novice and practiced drivers in this study, it could possibly be said that inclusion of a practiced group of drivers meant the current findings are perhaps a more accurate representation of the effects of different types of conversation across a broader range of road users.

One difference between the two studies that cannot be explained by methodological differences is the failure of the current study to observe pauses in participants' speech. Given that a lack of pauses was also observed in Experiment 1, which was a between-subjects design and did not include any practiced participants, the effects of the within-subjects design and inclusion of practiced participants must be ruled out. Coding and measurement of pauses were identical across the two studies; therefore the failure of the current experiments to observe pauses in participants' speech must be due to an unidentified factor.

It must also be acknowledged that there may have been changes in other aspects of speech that were not examined in this study. Rather than pausing or finishing an utterance, some participants chose to elongate certain words or use a greater number of fillers, such as "um," as they passed through hazards. While these both reflect changes in speech, they would not have been captured in the conversation measures utilised in this study. Laughter as a function was also not addressed. Mutual laughter serves to continue a conversation without thought needing to be given to processing or producing language, and as such may even serve the same purpose as a pause in some cases, allowing the driver time to concentrate on the driving situation without attending to the conversation.

Directions for future research

The ideas discussed in this section present a number of different avenues for future research. Someone intent on pursuing the role of visual cues in conversations may choose to incorporate additional features of a passenger conversation into a similar experiment, such as providing both cell phone and videophone conversers with information about the real-time driving situation and testing differences between these two remote conversation conditions. A researcher intrigued by workload theory or the impact of the nature of the driving task may design an experiment that systematically compares driving performance and workload over a range of driving hazards. The breadth of the findings uncovered in this study translates into a wide range of possible directions for any future research stemming from this study.

References

Alm, H., & Nilsson, L. (1994). Changes in driver behaviour as a function of handsfree mobile phones - a simulator study. *Accident Analysis and Prevention, 26,* 441-451.

Alm, H., & Nilsson, L. (1995). The effects of a mobile telephone task on driver behaviour in a car following situation. *Accident Analysis and Prevention, 27*, 707-715.

Amado, S., & Ulupinar, P. (2005). The effects of conversation on attention and peripheral detection: Is talking with a passenger and talking on a cell phone different? *Transportation Research Part F, 8*, 383-395.

Becic, E., Dell, G. S., Bock, K., Garnsey, S. M., Kubose, T., & Kramer, A. F. (2010). Driving impairs talking. *Psychonomic Bulletin & Review, 17*(1), 15-21.

Beede, K. E., & Kass, S. J. (2006). Engrossed in conversation: The impact of cell phones on simulated driving performance. *Accident Analysis and Prevention, 38,* 415-421.

Brookhuis, K. A., de Vries, G., & de Waard, D. (1991). The effects of mobile telephoning on driving performance. *Accident Analysis and Prevention, 23,* 309-316.

Brown, I. D., Tickner, A. H., & Simmonds, D. C. V. (1969). Interference between concurrent tasks of driving and telephoning. *Journal of Applied Psychology, 53,* 419-424.

Caird, J. K., Willness, C. R., Steel, P., & Scialfa, C. (2008). A meta-analysis of the effects of cell phones on driver performance. *Accident Analysis and Prevention, 40,* 1282-1293.

Charlton, S. G. (2009). Driving while conversing: Cell phones that distract and passengers who react. *Accident Analysis and Prevention, 41,* 160-173.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NY: Lawrence Erlbaum Associates.

Consiglio, W., Driscoll, P., Witte, M., & Berg, W. P. (2003). Effect of cellular telephone conversations and other potential interference on reaction time in a braking response. *Accident Analysis and Prevention, 35,* 495-500.

Cooper, J. M., & Strayer, D. L. (2008). Effects of simulator practice and real-world experience on cell-phone–related driver distraction. *Human Factors, 50,* 893-902.

Crundall, D., Bains, M., Chapman, P., & Underwood, G. (2005). Regulating conversation during driving: a problem for mobile telephones? *Transportation Research Part F, 8,* 197-211.

Deery, H. A. (1999). Hazard and risk perception among young novice drivers. *Journal of Safety Research, 30,* 225-236.

Drews, F. A., Pasupathi, M., & Strayer, D. L. (2008). Passenger and cell phone conversations in simulated driving. *Journal of Experimental Psychology: Applied, 14*, 392-400.

Endsley, M. R. (1995). Towards a theory of situation awareness in dynamic systems. *Human Factors, 37*, 32-64.

Fuller, R. (2005). Towards a general theory of driver behaviour. *Accident Analysis and Prevention, 37,* 461-472.

Fuller, R. & Santos, J. A. (2002). Psychology and the highway engineer. In R. Fuller & J. A. Santos (Eds.), *Human factors for highway engineers* (pp. 1-10). Oxford, UK: Elsevier Science.

Gregersen, N. P. (1996). Young drivers' overestimation of their own skill – an experiment on the relation between training strategy and skill. *Accident Analysis and Prevention,* 28, 243-250.

Gugerty, L., Rakauskas, M., & Brooks, J. (2004). Effects of remote and in-person verbal interactions on verbalization rates and attention to dynamic spatial scenes. *Accident Analysis and Prevention, 36*, 1029-1043.

Hancock, P. A., Lesch, M., & Simmons, L. (2003). The distraction effects of phone use during a crucial driving maneuver. *Accident Analysis and Prevention, 35*, 501-514.

Harbluk, J. L., Noy, Y. I., Trbovich, P. L., & Eizenman, M. (2007). An on-road assessment of cognitive distraction: Impacts on drivers' visual behavior and braking performance. *Accident Analysis and Prevention, 39,* 372-379.

Harré, N., Field, J., & Kirkwood, B. (1996). Gender differences and areas of common concern in the driving behaviors and attitudes of adolescents. *Journal of Safety Research, 27*, 163-173.

Horrey, W. J., Lesch, M. F., & Garabet, A. (2008). Assessing the awareness of performance decrements in distracted drivers. *Accident Analysis and Prevention, 40,* 675-682.

Horrey, W. J., & Wickens, C. D. (2006). Examining the impact of cell phone conversations on driving using meta-analytic techniques. *Human Factors, 48*, 196-205.

Hunton, J., & Rose, J. M. (2005). Cellular telephones and driving performance: the effects of attentional demands on motor vehicle crash risk. *Risk Analysis, 25*, 855-866.

Irwin, J. D., & Chekaluk, E. (2006, October). *Can you talk and drive safely at the same time? An examination of gender differences in driving performance of young drivers observed whilst conversing with a passenger or on a mobile phone.* Paper presented at the Australasian Road Safety Research Policing and Education Conference. Retrieved from http://hdl.handle.net/1959.14/7027

Jonah, B. A. (1990). Age differences in risky driving. *Health Education Research: Theory and Practice, 5,* 139-149.

McEvoy, S. P., Stevenson, M. R., McCartt, A. T., Woodward, M., Haworth, C., Palamara, P., & Cercarelli, R. (2005). Role of mobile phones in motor vehicle crashes resulting in hospital attendance: a case-crossover study. *BMJ, 331*, 428-430.

McEvoy, S. P., Stevenson, M. R., & Woodward, M. (2007). The contribution of passengers versus mobile phone use to motor vehicle crashes resulting in hospital attendance by the driver. *Accident Analysis and Prevention, 39*, 1170-1176.

McKnight, A. J., & McKnight, A. S. (1993). The effect of cellular phone use upon driver attention. *Accident Analysis and Prevention, 25*, 259-265.

Nunes, L., & Recarte, M. A. (2002). Cognitive demands of hands-free-phone conversation while driving. *Transportation Research Part F, 5*, 133-144.

Patten, C. J. D., Kircher, A., Östlund, J., & Nilsson, L. (2004). Using mobile telephones: cognitive workload and attention resource allocation. *Accident Analysis and Prevention, 36,* 341-350.

Rakauskas, M. E., Gugerty, L. J., & Ward, N. J. (2004). Effects of naturalistic cell phone conversations on driving performance. *Journal of Safety Research, 35*, 453-464.

Redelmeier, D. A., & Tibshirani, R. J. (1997). Association between cellular-telephone calls and motor vehicle collisions. *The New England Journal of Medicine, 336*, 453-458.

Rivardo, M. G., Pacella, M. L., & Klein, B. A. (2008). Simulated driving is worse with a passenger than a simulated cellular telephone converser. *North American Journal of Psychology, 10,* 265-276.

Rosenbloom, T. (2006). Driving performance while using cell phones: An observational study. *Journal of Safety Research, 37*, 207-212.

Shinar, D., Tractinsky, N., & Compton, R. (2005). Effects of practice, age, and task demands, on interference from a phone task while driving. *Accident Analysis and Prevention, 37*, 315-326.

Simons-Morton, B., Lerner, N., & Singer, J. (2005). The observed effects of teenage passengers on the risky driving behavior of teenage drivers. *Accident Analysis and Prevention, 37,* 973-982.

Strayer, D. L., Drews, F. A., & Crouch, D. J. (2006). A comparison of the cell phone driver and the drunk driver. *Human Factors, 48*, 381-391.

Strayer, D. L., Drews, F. A., & Johnston, W. A. (2003). Cell phone-induced failures of visual attention during simulated driving. *Journal of Experimental Psychology: Applied, 9,* 23-32.

Strayer, D. L., & Johnston, W. A. (2001). Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular telephone. *Psychological Science, 12*, 462-466.

Svenson, O., & Patten, C. J. D. (2005). Mobile phones and driving: a review of contemporary research. *Cognition, Technology & Work, 7*, 182-197.

Tornros, J. E. B., & Bolling, A. K. (2005). Mobile phone use - Effects of handheld and handsfree phones on driving performance. *Accident Analysis and Prevention, 37*, 902-909.

Tsimhoni, O., & Green, P. (2001). Visual demand of driving and the execution of display-intensive in-vehicle tasks. *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting, 2*, 1586-1590.

Waugh, J. D., Glumm, M. M., Kilduff, P. W., Tauson, R. A., Smyth, C. C., & Pillalamarri, R. S. (2000). Cognitive workload while driving and talking on a cellular phone or to a passenger. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 6*, 276-279.

Wood, J., Chaparro, A., Hickson, L., Thyer, N., Carter, P., Hancock, J., … Ybarzabal, F. (2006). The effect of auditory and visual distractors on the useful field of view: Implications for the driving task. *Investigative Ophthalmology & Visual Science, 47*, 4646-4650.

Appendices

Appendix A: Participant Information Sheet – Experiment 1

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

Participant Information Sheet

Videophone Effect Study

TARS
Traffic and Road Safety Research Group

The purpose of the study is to find out about the effects of different modes of conversation on driving performance, in particular, the effect of using a video phone whilst driving.

I am asking participants in the study to:

1. Answer a brief set of questions about their driving habits.

2. Decide who will drive and who will be the converser (you may both be asked to drive without conversing). The driver will drive a simulated road in the driving simulator while conversing with their partner. The conversation may involve the partner sitting in the car as a passenger, or talking with the driver via a cell phone or video phone. The road is 25km long and involves traffic and hazards such as what you might expect on a typical stretch of NZ road. For your participation each of you will receive your choice of a $5 grocery voucher, a coffee voucher from Momento cafe, or a 1% course credit if you are taking a first year psychology course. You will be able to practise driving in the simulator before you begin. If at any stage the simulator makes you feel queasy or dizzy please let me know and I will stop the simulator immediately. Please do not be concerned if this happens. Some people just react this way to simulators. You will still both get your choice of thank you present for participating.

3. Following the drive I will ask each of you to answer a few questions about how you found the task

All information collected from you will remain anonymous and if you have any questions please ask. You can withdraw from the experiment at any time. The study has received ethics approval from the School of Psychology Ethics Committee.

If you are happy to participate I will now have you complete an informed consent form and then fill in a quick questionnaire about your driving habits.

Thank you in advance for your participation. *Kathy Mackenzie*

Appendix B: Habit questionnaire used in Experiments 1 and 2

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Habits Questionnaire

(all information provided will remain anonymous)

TARS
Traffic and Road Safety Research Group

Please circle, tick, or fill in the blanks where appropriate

Do you have a full license? YES NO

If yes, please continue the survey. If no, please stop immediately and inform the researcher

How long have you been a licensed driver? (since you passed your learner's test)

_________ years & _________ months

What kind of vehicle do you drive most often? (tick one)

- ❑ Motorbike
- ❑ Car
- ❑ Van or ute
- ❑ Truck
- ❑ Other _____________

How many kilometres do you drive in an average week? _______________ km

In the past year, have you been involved in any motor vehicle crashes? YES NO

If yes, how many? _____________

In the past year, have you received any driving infringements (including speed camera fines)?

YES NO

If yes, how many? ________

Do you own a cell phone?

YES NO

Before the new law was introduced, did you use a cell phone while driving?

YES NO

If yes, what for?

- ❑ Talking
- ❑ Texting
- ❑ Both

Have you used a video phone or video chat programme (e.g. Skype) before?

YES NO

If yes, how often do you use it on average?

- ❑ Every day
- ❑ Several times a week
- ❑ Once a week
- ❑ Once a fortnight
- ❑ Once a month
- ❑ Less than once a month

What is your age? _______________

What is your gender? Male Female

What is your ethnicity?

- ❑ Maori
- ❑ NZ European/Pakeha
- ❑ Other (please state)_____________

That is the end of the survey – Thank you very much for your answers.
Let the researcher know that you are finished and they will show you how to begin your session on the driving simulator. If you have any questions, please ask!

Appendix C: Participant instructions – Experiment 1

Participant Instructions for Driver

Driving in the simulator – Overview

The simulator drives just like a car with automatic transmission, except you do not need to put it into gear to take off. Use the accelerator and brake to control the vehicle's speed. The car has a working speedometer on the dashboard so use this to monitor your speed. You control the vehicle's direction, but please keep to the main road. Please don't turn down any side streets. In some ways, driving the simulator is similar to driving in wet conditions. Avoid braking and changing direction at the same time. Please also be aware the suggested curve speeds are accurate. If you attempt to take a 65km/h corner at 110km/h you are likely to spin out. To get used to driving the simulator, you'll be taken through some training exercises and be given time to practice (about 10-15 minutes).

The experimental drive

After your practice you will drive a 25km section of road. This will take approximately 15-20 mins. Please treat today's drive as if you were driving your own car on an actual road e.g. if you normally drive slightly over or under the speed limit please do that today (speed limit signs are posted at various points). There are other vehicles on the road, so if you would normally pass them in an actual driving situation please do so here. Please keep driving until you come to a sign-posted rest area, which you must pull into.

The conversation

If your drive involves talking with a partner today, feel free to talk about whatever you wish. However, you must start by completing one of the conversational tasks (your partner will be given a list). Remember your conversation will be recorded so please refrain from talking about sensitive topics you are not happy for me to hear about. When you are ready I will ask you to begin the conversation. A few minutes later I will ask you to start driving. Please continue the conversation until the end of the drive. If you run out of things to talk about try completing the other conversational tasks on the card.

Afterwards

Upon completion you will be asked to watch a video of your drive and assess your level of mental workload at each stage of the drive. You will also be asked a few quick questions about the drive and conversation.

Do you have any questions before we begin?

Participant Instructions for Converser

During your conversation today, feel free to talk about whatever you wish. However, you must begin by completing one of the conversational tasks from the card given to you. Your conversation will begin a couple of minutes before your partner begins driving. Please keep the conversation going until the drive is complete. (If you are in the cell phone or video phone condition, you will be signaled when the drive is over.) If you run out of things to talk about, try completing the other conversational tasks on the card. Remember your conversation will be recorded so please refrain from speaking about sensitive topics you are not happy for me to hear about. Following the drive, you will be asked a few quick questions about the conversation.

Do you have any questions before we begin?

Appendix D: Conversation ideas card used in Experiments 1 and 2

CONVERSATION IDEAS

During today's drive, you are free to talk about what you wish. If you run out of things to talk about, try carrying out one of the following tasks:

Option 1:

- Come up with 12 items you both agree to take to a deserted island with you. Assume you will be on the island for two weeks before being rescued.

Option 2:

- Come up with 12 songs you both agree to put on a mix tape for a road trip.

Option 3:

- Come up with 7 home-cooked dinners you would both be happy to eat over the next week, and create a grocery list for all the necessary ingredients. Aim to stay within a budget of $120.

If your conversation naturally progresses onto other topics while completing this task, feel free to move on. Just ensure your conversation continues throughout the drive.

Appendix E: Participant information sheet – Experiment 2

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Videophone Effect Study Information Sheet

TARS
Traffic and Road Safety Research Group

The purpose of the study is to find out about the effects of different modes of conversation on driving performance, in particular, the effect of using a video phone whilst driving.

I am asking participants in the study to:

1) **Answer a brief set of questions about their driving habits.**

2) **Drive a short simulated road in the driving simulator four times, each time under different conversational conditions. One participant will be the driver and the other the conversation partner. The four conditions include a no-conversation control; passenger conversation; cell phone conversation; and videophone conversation.**

3) **Provide difficulty and interference ratings at the conclusion of each drive**

- The road is 10km long and involves traffic and hazards such as what you might expect on a typical stretch of NZ road. You will be given a short practice drive in the simulator before you begin.

- For your participation each of you will receive a $10 voucher.

- If at any stage the simulator makes you feel queasy or dizzy, please don't hesitate to let me know and I will stop the simulator immediately. Please do not be concerned if this happens, some people just react this way to simulators. It is better for your comfort and my research for you to stop if you feel ill rather than pushing through it. You will still both get the voucher for participating.

- All information collected during the experiment will remain anonymous. You can withdraw from the experiment at any time. The study has received ethics approval from the Dept. Psychology Ethics Committee.

If you are happy to participate I will now have you complete an informed consent form and then fill in a quick questionnaire about your driving habits. If you have any questions please ask. Thank you for your participation.

Kathy Mackenzie

Appendix F: Participant instructions – Experiment 2

Participant instructions for driver

Many of the driving participants in this experiment are well-practiced in the simulator; however, as some time may have passed since you last drove it, you will be given a short practice drive to reacquaint yourself with the simulator and to familiarise yourself with the road to be used in this experiment. If you are a novice driver, you will be taught how to use the simulator then will complete the same practice drive.

After your practice you will drive a 10km section of road four times. Each drive will take approximately 8mins. At the end of this time you will see a sign-posted rest area on the left hand side. Please finish your drive by pulling into it. Please treat today's drive as if you were driving your own car on an actual road e.g. if you normally drive slightly over or under the speed limit please do that today (speed limit signs are posted at various points). There are other vehicles on the road, so if you would normally pass them in an actual driving situation please do so here.

Each of your four experimental drives will involve a different conversational condition – a no-conversation control, passenger, cell phone or videophone. The experimenter will inform you of which condition you will do each time and will help you get set up for it. Throughout the experiment, you are free to talk about whatever you wish. Please keep the conversation going throughout each drive, but try to let it flow naturally. Your partner will be given a card of conversation ideas to use if you run out of things to talk about.

Please be aware your conversation will be recorded so refrain from talking about sensitive topics you are not happy for me or the experimenter to overhear. Aspects of your driving performance (e.g. speed, braking etc.) will be recorded throughout the drive. Following each drive you will be asked to give difficulty and interference ratings.

Do you have any questions before we begin?

Participant instructions for converser

During today's experiment, your partner will drive the same road four times. Each of the four experimental drives will involve a different conversational condition – a no-conversation control, passenger, cell phone or videophone. You will act as the conversation partner for each of the conversation conditions. The experimenter will inform you of which condition you will do each time and will help you get set up for it.

During the experiment you are free to talk with your partner about whatever you wish. Please keep the conversation going until the end of each drive (when you are in the cell phone or videophone condition, you will be signalled when the drive is over). Try to let the conversation flow as naturally as possible. You will be given a card of conversation ideas to use if you run out of things to talk about.

Please be aware your conversation will be recorded so please refrain from speaking about sensitive topics you are not happy for me or the experimenter to overhear. Following each drive, you will be asked to rate the conversation in terms of difficulty and interference.

Do you have any questions before we begin?

Appendix G: Table showing counterbalanced order for participant trials used in Experiment 2

Order for participant trials

	Trial 1	Trial 2	Trial 3	Trial 4
Participant 1	Control	Passenger	Cell phone	Videophone
Participant 2	Passenger	Control	Videophone	Cell phone
Participant 3	Cell phone	Videophone	Control	Passenger
Participant 4	Videophone	Cell phone	Passenger	Control
Participant 5	Control	Videophone	Passenger	Cell phone
Participant 6	Passenger	Cell phone	Control	Videophone
Participant 7	Cell phone	Passenger	Videophone	Control
Participant 8	Videophone	Control	Cell phone	Passenger
Participant 9	Control	Cell phone	Videophone	Passenger
Participant 10	Passenger	Videophone	Cell phone	Control
Participant 11	Cell phone	Control	Passenger	Videophone
Participant 12	Videophone	Passenger	Control	Cell phone